Supporting Information

Structural Transition in Physical Networks

Nima Dehmamy,[1] Soodabeh Milanlouei,[1] and Albert-László Barabási[1,2,3]

[1]*Network Science Institute, Center for Complex Network Research, and Department of Physics, Northeastern University, Boston, USA*

[2]*Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, USA*

[3]*Department of Network and Data Sciences, Central European University, Budapest, Hungary*

(Dated: August 15, 2018)

CONTENTS

SI 1. INTRODUCTION

Some networks, such as the network of neurons in the brain, reside in three-dimensional space. Links and nodes in these networks occupy space. Many aspects of such networks are studied as part of "gelation" processes in polymer physics (see [19] chapter IV for instance). Gels are like frozen polymers with inter-linking, like vulcanized rubber. But in polymers, the ideal situation with maximally stretched links (occupying minimal volume) is not generally of interest, as it is highly unnatural. For problems such as the organization of the white matter of the brain, however, each long fiber has a cost to produce and thus the system must be very economical.

To understand the structural properties of networks in 3D, we must understand the constraints that being embedded in 3D puts on the network. In particular, the layout algorithm used to embed the network in 3D must take the thickness of the links and node sizes into account. Most layout algorithms used today are derived from the "Force-directed Layout" (FDL) algorithms [20, 21]. FDL has the benefit that nodes connected to each other are pulled

closer and, thus, it clusters nodes that belong to the same network modules [22]. However, FDL without modification does not take link thickness into account. Here we develop a layout algorithm which is based on FDL, but which allows for efficient layout of networks in 3D, helping us understand the physical properties of such layouts. By efficient layout we are referring to a layout which aims to minimize the space occupied by the nodes and links. In other words, the algorithm tries to find the layout with the shortest links, while avoiding crossings between links or nodes. We achieve this by defining repulsive forces among links and among nodes. As we will discuss below, finding the real optimum of this problem is NP-hard. However, using heuristic methods, such as simulated annealing, allows us to find reasonably good layouts in a short time.

Our main result that networks in 3D will have two main phases: *Weakly interacting regime*, where links are very thin, compared to node sizes, and link crossing is rare; *Strongly interacting regime*, where links are thick, filling the space, and have to bend and curve significantly to avoid crossing. The existence of these two phases is observed in various order parameters. We use the average link length, average link curvature, as well as the relaxation time for simulations to quantify these two phases and to identify where the transition occurs. The details of the definition of link curvature and how we define the relaxation time are given in sec. SI 5.

SI 2. LINK CROSSING IN LAYOUTS WITH STRAIGHT LINKS

Before examining the bent links, we first discuss the behavior of link crossing in 3D embedded networks when the links are straight. Analyzing the number of crossings as a function of link thickness in FDL yields a linearly growing number of crossings (Fig. SI.4.) The point where crossings become noticeable depends mostly on the network density parameters, similar to how in self-avoiding polymers the excluded volume depends on the interaction strength. Figure SI.5 shows that as we increase the link thickness, the number of crossings grows linearly. We show below that randomly oriented cylindrical links will result in a linear trend for the number of crossings versus link thickness.

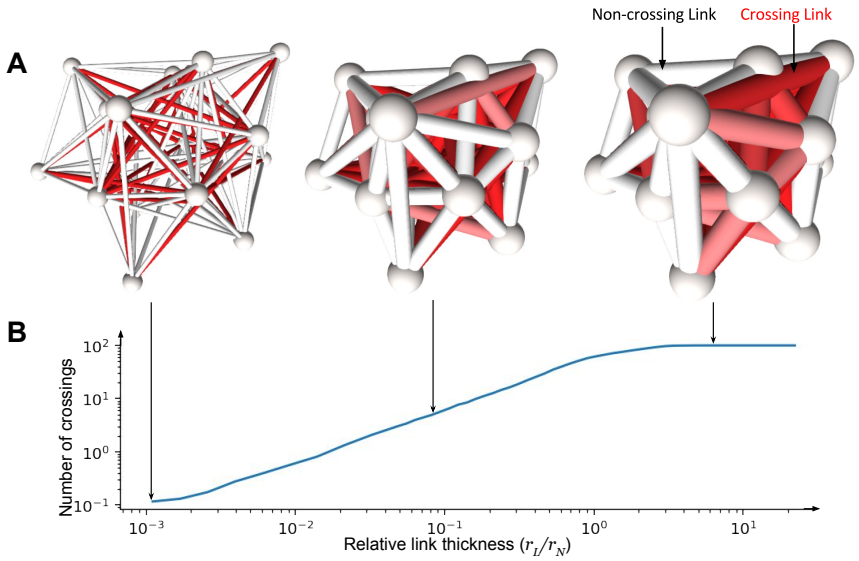

FIG. SI.4. **(A)** Crossing links for a Barabasi-Albert (BA) network ($N = 20, m = 3$) in space, drawn by FDL. While increasing the link thickness r_L, the network occupies more space and crossings increase. **(B)** The plot shows the trend by which the number of crossings increases in a BA network with $N = 100, m = 3$ laid out in 3D using FDL. The number of crossings increases linearly with r_L at first, before tapering off due to finiteness of the network. It can be shown (SI 2) that layouts which result in random link orientations yield such linear trend in the growth of the number of crossings.

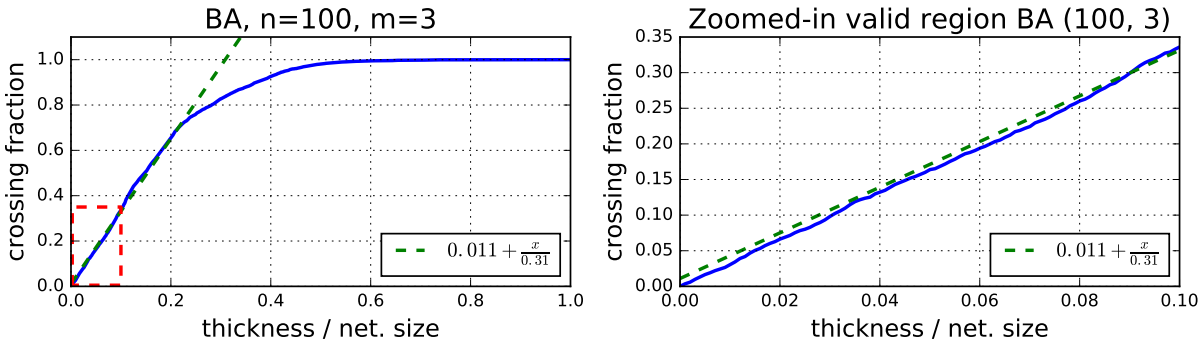

FIG. SI.5. Left: The behavior of the number of crossings as a function of link thickness. Trivially, once the thickness becomes comparable to the network size (when "thickness/net. size" becomes order 1) the crossings saturate and most links cross. The region in which trying to make links avoid each other is much smaller. It extends from zero thickness up to where thickness / net. size is of order $1/\sqrt{N_e}$ with N_e being the number of links in the network. In the BA network with 100 nodes and $m = 3$ considered here $1/\sqrt{N_3} \approx 0.06$, which is a safe lower limit for the region where avoiding crossings is possible. Right: a zoomed-in version of the dashed-red box on the lest highlighting this region of validity of avoiding crossings. The legends show a good linear fit to the data up to thickness / net. size = 0.2.

A. Number of Conflicts: Phase Diagram

We wish to find the number of crossings as a function of link radius. We will divide this radius by $N^{1/3}$ to get the average space per node. The number of links should be of order $2N$ because each new node brings in exactly two new links. Thus, we expect most links to intersect once their radius becomes of order $r/2$. When the thickness is very small, we expect

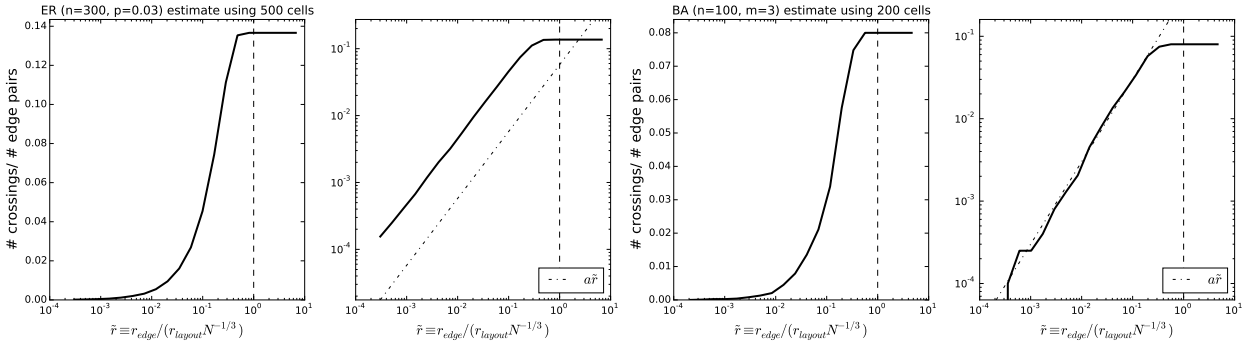

FIG. SI.6. Simulations of number of crossings in BA and ER networks, when laid out using FDL. When plotted in log-log, we see that both networks show a linear trend in the growth of the number of link conflicts as a function of the link thickness. To efficiently calculate the number of crossings, the links are broken into small segment and the space is partitioned into cells. Then, only conflicts within cells with more than one link crossing them are counted. This avoids order L^2 computation.

there to be very few crossings. In this case, perturbing the layout should be able to resolve the crossing. At very high thickness, though we know that the layout may get clogged up so that no perturbation could resolve the crossings. There should be a qualitative difference between these two states.

Simulations of Barabasi-Albert and Erdös-Renyi networks layed out in 3D using the Fruchterman-Reingold forced-directed layout seems to suggest that in this setting the qualitative behavior of these networks for different values of parameters is very similar. Figures SI.7, SI.8 and SI.9 show simulations with parallel, randomly oriented and finite links, respectively. The plateau in the number of crossings at large radii is due to the limitation of the optimized algorithm, which partitions the space into smaller boxes and fails when the link radius reaches box size. As we can see in Figure SI.6, the number of crossings in both types of networks and in all four simulations grows linearly with link radius. The question is whether this is due to the layouts randomness.

B. Random Crossings

What is the source of the linear growth of the number of crossings between links with link thickness? We show here that it depends a lot on the way the links are arranged in space. In particular, we first show that if the links were arranged parallel to each other, but at random distances from each other, the number of crossings among them would grow quadratically with the thickness, inconsistent with the observed trend in FDL. Next, we

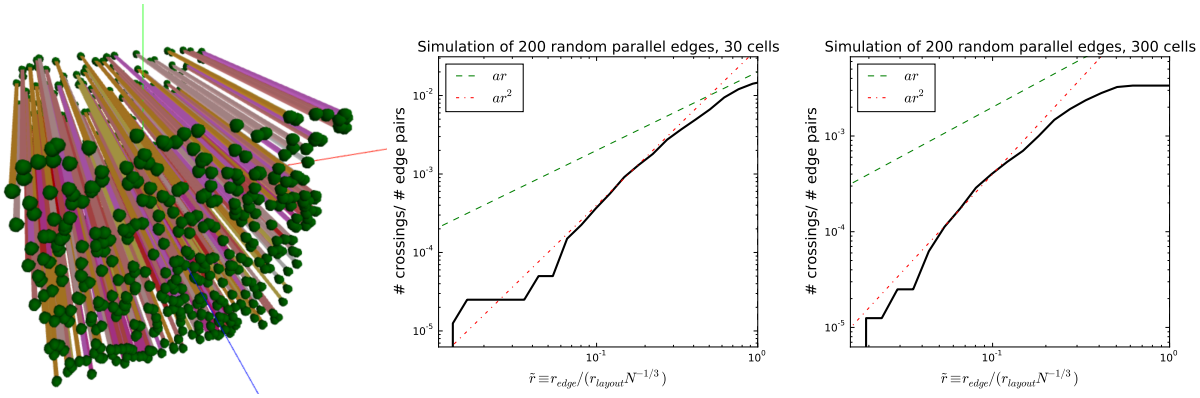

FIG. SI.7. Simulated crossing count of 200 parallel links. One simulation used 30 partition cells in space and the other used 300 cells. In both the number of crossings grows as r^2, in accordance with the theoretical expectation. This is in contrast with the linear trend observed in the link crossing of force-directed layouts of random graphs.

show that the observed linear trend is reproducible if the links are either randomly oriented, or even grouped in two groups of parallel links but in two different spatial dimensions (e.g. one stack parallel to the x-axis, another to the y-axis).

1. *Parallel Links with Random Distances*

For simplicity, let's first consider a number of dots scattered with a random uniform distribution in d dimensional space with a density ρ. Suppose we want to grow d dimensional balls around these dots and wish to count the number of balls that overlap. Because of the uniformity, the total number of balls of radius r crossing each other is the number of points that are within less than $2r$ distance of each other, which is

$$n(r) = \int_0^{2r} \rho(r) dVol = \rho \Omega_{d-1} (2r)^d \tag{SI.5}$$

which scales with the volume of space. In two dimensions, for example, this would mean that the number of circles of radius r crossing each other will be $4\pi\rho r^2$. To compare with the problem of link crossing, if we had parallel cylindrical links in 3D, their cross-section would have been randomly scattered circles in 2D and again the number of crossings should grow quadratically with radius.

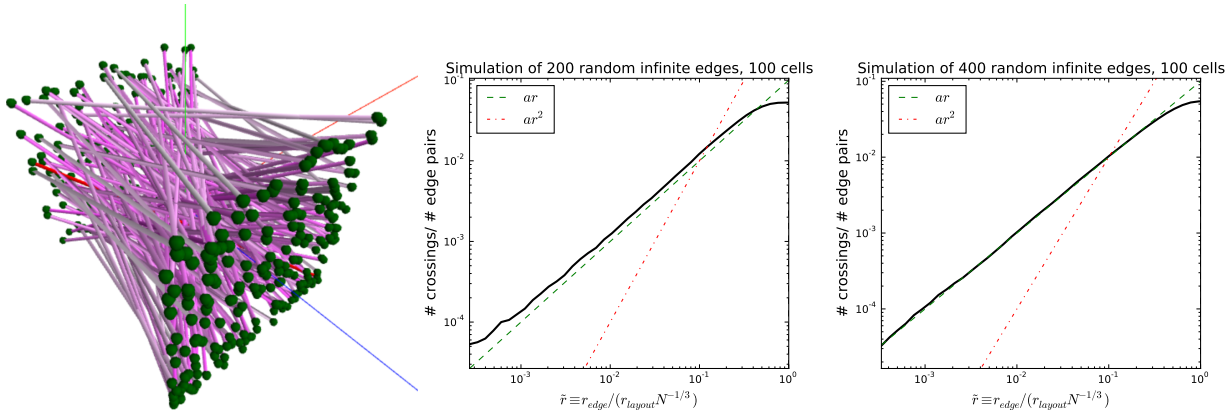

FIG. SI.8. Simulated crossing count of randomly oriented links. In both the number of crossings grows as linearly with r, in contrast to the parallel link case. This is the same trend as the one observed in the link crossing of force-directed layouts of random graphs.

2. Randomly Oriented Links

Now consider infinite cylindrical links scattered with random orientations in a 3D space. Assume that the space occupied by the cylinders has a roughly constant density within cubes of links $d \gg r$. Take the cross-section of this space with a flat plane. The center lines of the cylinders are uniformly distributed points on the plane and the cross-section of the cylinders are ellipses. The ellipses may be very eccentric, but the minor axis, which is also the radius of the largest circle that fits inside the ellipse, is again r. Thus, regardless of the eccentricity, if two dots on the plane (the cross-section of center lines of two links) are closer than $2r$ from each other, then the two links will definitely cross. Again, the number of dots less than $2r$ away from each other on any random flat plane grows with r^2 and this is only a lower limit for the crossings assuming zero eccentricity for the cross-sectional ellipses. Therefore the number of link crossings should also grow at least quadratically in 3D.

3. Two and Three Stacks of Parallel Links along Different Axes

However, contrary to this argument, as Fig. SI.8 shows, the number of crossings among randomly oriented infinite links does grow only linearly with radius, as it did in the random graph layouts. The same linear trend is observed in finite random links.

But there are other simpler examples which also show a linear trend in the growth of the number of collisions. For instance, if we take a number of random parallel links and add a

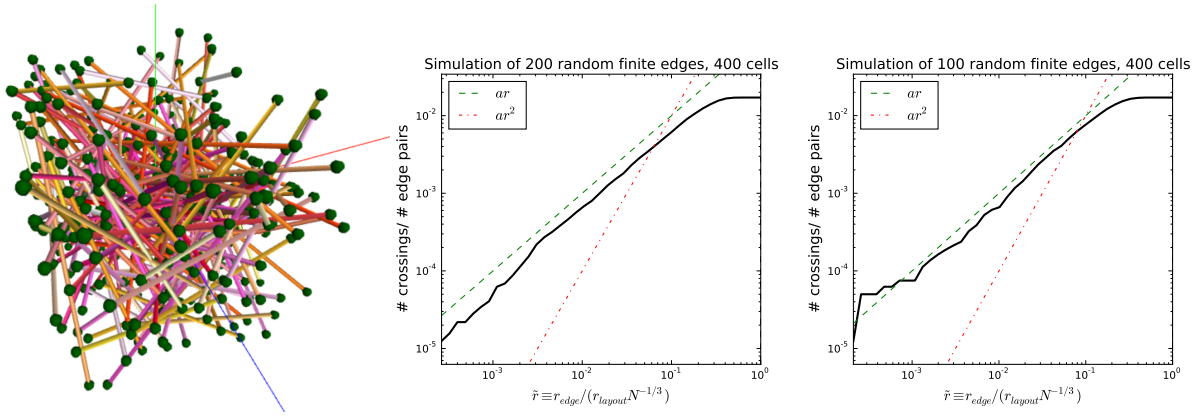

FIG. SI.9. Simulated crossing count of randomly oriented finite links which shows the same linear trend as the infinite randomly oriented links, suggesting that the random orientation is cause for the trend.

similar number of links that go in a direction perpendicular to those links the trend becomes linear instead of quadratic, as is demonstrated in Figure SI.10. The trend is linear for two sets of parallel links, where one parallel set is perpendicular to the other parallel set. The same is true for three perpendicular sets of parallel links.

C. Deriving the Linear Growth of Conflicts

The number of links within distance r of a point on link a can be found by projecting all links onto the plane to which link a is normal. Since the angle distribution is random, we get a uniform distribution of lines on this plane and the number of lines within distance r of a point on a 2D plane grows linear in r (Fig. SI.11.)

Consider a set of infinite parallel lines distributed uniformly in 2D with linear density λ in the direction perpendicular to the lines. Now consider a point p on this plane. The number of links that a circle of radius r centered at p crosses is simply

$$n(r) = \lambda r$$

and is a linear function of r because only the projection in the direction perpendicular to the lines determines whether or not they cross the circle.

Now, consider n random lines in a 2D plane with a uniform density (i.e. number of lines passing through a unit square) λ. Again, the number of lines passing a circle of radius r around a point p grows linearly with r. To see this, take a line l. Draw the line that passes

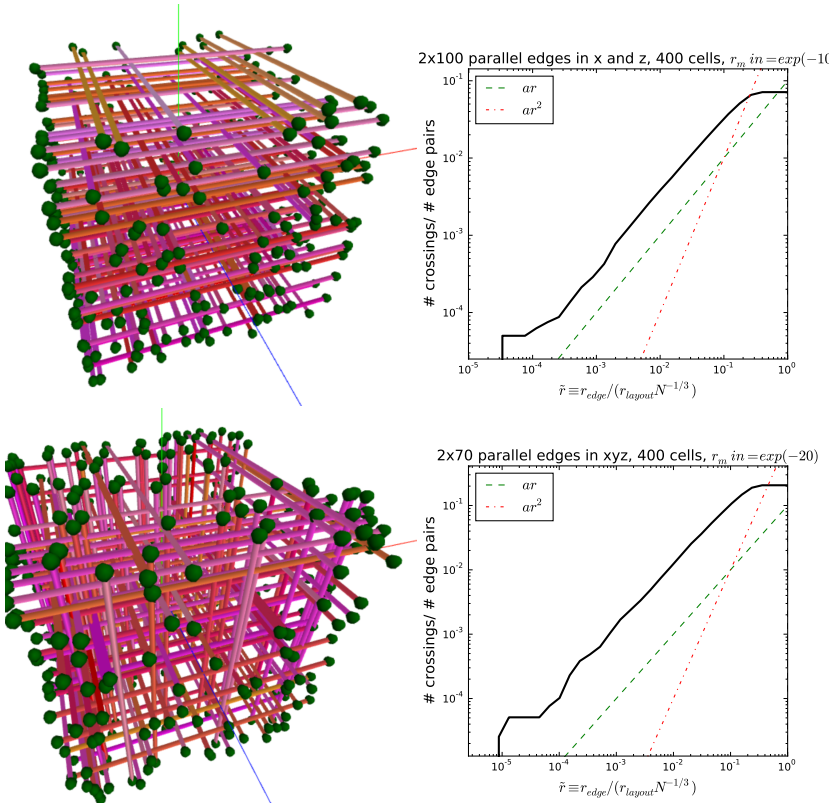

FIG. SI.10. Simulated crossing count of 2 and 3 sets parallel links with each set running perpendicular to the other sets. The trend of the number of crossings becomes linear, which is different from the quadratic trend observed in oe set of parallel links in Fig. SI.7 and agrees with the linear trend observed in the link crossing of force-directed layouts of random graphs and in the randomly oriented links in Figs. SI.8 and SI.9, suggesting that being having components perpendicular to another link may be the key to the linear trend.

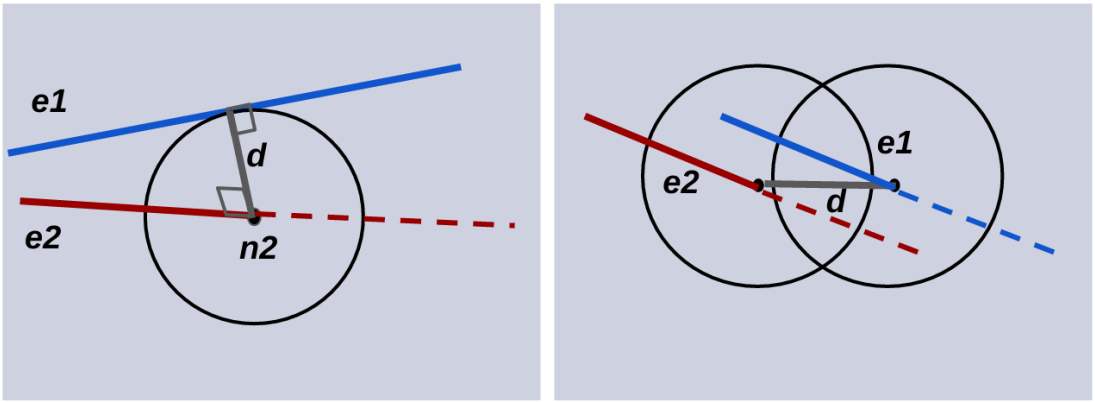

FIG. SI.11. **Left:** The projection of the line through the center of an link **e1** on the plane normal to link **e2** which crosses the link at point **n2**. Because links are randomly oriented, the distance of **e1** and **e2** is the distance of the blue line from the point **n2**. The uniformity then means that the blue lines have uniform density at random angles on the plane and thus their number within distance d is linear in d. **Right:** If links were parallel, the projection of **e1** onto the plane would be a point. Thus the distance of the links is the distance of two points on the plane. Uniform spatial distribution in this case means we get uniformly distributed points on the plane and their number within a circle of radius d grows as d^2.

through p and lands perpendicular to l. Say the distance from p to l is d. This distance determines whether or not l crosses a circle around p. Because of the rotational symmetry, if we rotate all lines around p and make all lines parallel to each other, we will not change the distribution of their distances from p. Since they were distributed uniformly in space, the parallel version will also have a uniform distribution. As we showed above, the number of crossings for the uniformly distributed lines grows linearly with r and thus so does the trend for randomly oriented lines. The point p above can be thought of as the center of the cross-section of a link running perpendicular to a plane containing other links. Thus, the number of crossings of a link that is perpendicular to a number of other links inside a plane, either parallel links or randomly oriented ones, will grow linearly with the radius of the cross-section of links. To see this, take a plane crossing a link (with the link coming out normal to it). Project all the other links onto this plane. The projections of the centroid of these links will be randomly oriented links on this plane. Now, any line that is within distance $2r$ of p results in a conflict with our reference link. If the links are uniformly, randomly distributed in space, the number of links within $2r$ grows linearly with r, as argued above. This argument only fails if the projections of the other links is not a line in this plane, which only happens if all links are parallel to the reference link. We already showed that in that case the crossings grow quadratically.

Thus, having links that have components in at least two perpendicular directions in 3D leads to a linear trend in the number of crossings as a function of r.

D. Finding Overlapping Links

Consider two straight links A and B. Denote the two vectors pointing to the endpoints of the two links by A_1, A_2 and B_1, B_2. Define the $a = A_1 - A_2$ and $b = B_1 - B_2$ which are parallel to the lines connecting the two endpoints of each link. The links are cylinders of diameter d. We wish to find whether or not the two links intersect, i.e. whether the two links come closer than a distance d of each other. The shortest distance between the two vectors is $a \times b$, since this is perpendicular to both a and b. Define

$$c = a \times b$$

we can decompose a to its components parallel and transverse to b

$$\begin{aligned} a &= a_{\parallel}\hat{b} + a_{\perp}(\hat{b} \times \hat{c}) \\ a_{\parallel} &= a \cdot \hat{b}, \\ a_{\perp} &= a \cdot (\hat{b} \times \hat{c}) = \hat{c} \cdot (a \times \hat{b}) = \frac{|c|}{|b|} \end{aligned} \tag{SI.6}$$

1. Measuring Link Distance

Consider the plane P spanned by b and c and containing the link B. The unit vectors $\hat{b}$ and $\hat{c}$ form an orthonormal coordinate system for this plane and any point on the plane can be written as

$$x_p = x_m\hat{b} + x_c\hat{c}$$

The plane itself is defined through the normal vector

$$\hat{n} = \hat{b} \times \hat{c}$$

and any multiple $p\hat{n}$ of $\hat{n}$ from the origin intersects the plane. When a and b are not parallel to each other, a will cross the plane P at a single point. The projection of link A onto P is a line parallel to B. Therefore we can use the following to find the distance between the two links. We will first pick one endpoint from each of A and B. The vector connecting the two

$$d = A_1 - B_1$$

is a vector whose projection of the vector $c = a \times b$ is perpendicular to both a and b and thus also to the projection of a onto P. Since on P the projection of a is parallel to b, the magnitude of the projection d_p of d onto P is the minimum distance between the two links. Thus we have

$$|d_p| = d \cdot \hat{c}$$

2. Caveat

The above-mentioned d_p is the least distance between the two links only if A crosses the plane P. Otherwise, one of the endpoints will be the closest point to the other line and we need a different strategy. Let's first figure out a way to see if A crosses P or not. To this aim we will look at the projections of $d_1 = A_1 - B_1$ and $d_2 = A_2 - B_2$ along the normal $\hat{n}$. Regardless of which endpoints of A are being connected to which on B, if the projections of $d_1 \cdot \hat{n}$ and $d_2 \cdot \hat{n}$ have the same sign it means A did not cross P and if they have opposite signs or one is zero A will have crossed P. Thus

$$\text{if: } (d_1 \cdot \hat{n})(d_2 \cdot \hat{n}) > 0 \quad \Rightarrow A \text{ not crossing } P$$

And so the closest point on A to B is one of A's endpoints. Similarly, we need to check if B crosses the plane spanned by A and $\hat{c}$. If not, the closest distance of the A and B will be one of their endpoints to each other.

SI 3. ELASTIC LINK MODELS: ELI AND FUEL

Our main goal is to understand the volumetric constraints on embedding networks in 3D. To achieve this, we propose a simple model that minimizes the length of links, while avoiding link crossing. We will begin by deriving mathematically how to find the shortest path between two points, in the presence of obstacles, showing that it is equivalent to stretching a rubber band between them. Next, we define smooth alternatives to hard-core repulsion to define repulsive forces among links and among nodes to make crossings energetically unfavorable. We then combine all the potential energy terms and derive the equations of motion in a strongly dissipative medium, to make sure the system moves to a local minimum energy state.

A. Equivalence of Shortest Path and Stretched Rubber Band

To find a globally optimum geodesic, be it in flat or curved space, one has to check many different paths and locally optimize them. A locally optimized path, meaning every

segment is as short as possible, is a geodesic. But in curved space there exist many geodesics connecting two points. This is because the geodesic equation is a second order ODE and different initial velocities may exist which make the geodesic cross the same final point in space. Denote the components of the metric of the space by $g_{\mu\nu}$. Heuristically, since a geodesic is locally the shortest path, a simple way to find geodesics between points should be to assume that they are stretched rubber bands. This assertion can be made rigorous. The key point is that for a geodesic we are minimizing the total path length locally, which means that we are minimizing the action

$$S[\gamma; a, b] = \int_\gamma dt \sqrt{g_{\mu\nu}(x) \frac{dx^\mu(t)}{dt} \frac{dx^\nu(t)}{dt}} \tag{SI.7}$$

where γ denotes the geodesic path. The energy for a rubber band has a similar form, but it lacks the square root

$$E[\gamma; a, b] = \frac{k}{2} \int_\gamma g_{\mu\nu}(x) \frac{dx^\mu(t)}{dt} \frac{dx^\nu(t)}{dt} dt \tag{SI.8}$$

But if the parameter t is itself an "affine parameter" (e.g. using the actual proper length element $dl = dS$ as the parameter, with S being the action above), we can show that the equations of motion derived from S and E become the same. To see this, we will derive the Euler-Lagrange equations from varying S and E. First define

$$dl \equiv \sqrt{g_{\mu\nu}(x) dx^\mu(t) dx^\nu(t)},$$

and for convenience $f(x) \equiv kx^2/2$ so that $E = \int f(dl/dt)dt$. Varying S yields

$$\delta S = \int dt \delta \left(\frac{dl}{dt} \right) = \frac{\partial \dot{l}}{\partial \dot{x}^\mu} \delta \dot{x}^\mu \Big|_{t_0}^{t_f} - \int dt \delta x^\mu \frac{d}{dt} \left(\frac{\partial \dot{l}}{\partial \dot{x}^\mu} \right) \tag{SI.9}$$

where the first term is a boundary term and vanishes because $\delta\dot{x}^\mu = 0$ at the boundaries. From varying E we have

$$\begin{aligned}\delta E &= \int dt \frac{df(\dot{l})}{d\dot{l}} \frac{\delta\dot{l}}{\delta x^\mu} \delta x^\mu \\ &= \frac{df(\dot{l})}{d\dot{l}} \frac{\partial\dot{l}}{\partial\dot{x}^\mu} \delta\dot{x}^\mu \Big|_{t_0}^{t_f} - \int dt \delta x^\mu \frac{d}{dt}\left(\frac{df(\dot{l})}{d\dot{l}} \frac{\partial\dot{l}}{\partial\dot{x}^\mu}\right) \\ &= -\int dt \delta x^\mu \left[\frac{\partial\dot{l}}{\partial\dot{x}^\mu} \frac{d}{dt}\left(\frac{df(\dot{l})}{d\dot{l}}\right) + \frac{df(\dot{l})}{d\dot{l}} \frac{d}{dt}\left(\frac{\partial\dot{l}}{\partial\dot{x}^\mu}\right)\right] \end{aligned} \tag{SI.10}$$

In the last line, if we only had the second term minimizing E would require the same equation as minimizing S. The first term here contains

$$\frac{d}{dt}\left(\frac{df(\dot{l})}{d\dot{l}}\right) = k\frac{d\dot{l}}{dt} = \frac{d^2 l}{dt^2}$$

If the parameter t is an affine parameter, say $t \propto l$, i.e. proportional to the proper length l itself, this term is proportional to $d^2l/dl^2 = 0$ and would vanishes. So, an affinely parametrized geodesic and the energy function of a rubber band with the same parametrization yield the same equations of motion. Thus, we can replace the problem of finding geodesics between two points with the problem of finding minimum energy configurations of a stretched rubber band connecting the two points. The difference in initial velocities of an object following the geodesic, which leads to having multiple geodesics connecting two points, is related to having different spring constants k.

B. Elastic Links

As is known about geodesics [1], and discussed in appendix SI 3 A, the problem of minimizing the length of links can be mapped to stretching elastic rubber bands between points in space[1]. The contribution of the stretched rubber band model of link l to the total energy is

$$E_l[\gamma_l] = \frac{k}{2}\int_{\gamma_l} ds_l \frac{d\vec{x}_l}{ds_l} \cdot \frac{d\vec{x}_l}{ds_l}, \tag{SI.11}$$

[1] The trade-off is that we lose re-parametrization invariance. The energy equation has a fixed parametrization ds_l which is an affine parametrization of the path.

where $d\vec{x}_l/ds_l$ denotes the tangent vector to link l (along the line through the center of the tube), γ_l is the path of the link, and k is the spring constant. The energy is obviously a function of the path γ_l. The full elastic potential V_{el} of the system is

$$V_{el} = \sum_l E_l \tag{SI.12}$$

The rubber links will occupy some space and exert a strong repulsive force on each other. Within one link, however, we will only assume no repulsive interactions. This allows the links to shrink to their minimal length under the elastic force of the stretched rubber band model and assume the equilibrium length of the band the be zero to allow for maximum shrinking.

C. Repulsive Potential

For large molecules, one usually uses a potential ansatz like the Lennard-Jones potential $V \sim (r_L/r)^{12} - (r_L/r)^6$ to simulate the van der Waals interactions, which is weakly attractive at large distances and strongly repulsive at short distances of order r_m. Since we are not concerned with molecules here but rather spatially embedded networks in general, we will not assume any attractive force between links and only assume there is a strong repulsion once link walls get close to each other. The simplest way to model this would be a repulsive force $F = -\nabla V$ that grows quickly once links get closer than twice the link thickness. In the ideal case, the force is infinite when the two touch or have distance less than twice the thickness. But a force coming from an infinite potential does not behave well in numerical simulations. A remedy is to simply not allow segments to move into each other by fixing minimum distance. But a better strategy numerically is to "regularize" the potential by replacing it with a finite potential that achieves approximately the same results.

The simplest regularization would be to replace the wall potential with a steep incline

$$V \sim \begin{cases} -A(r - r_L) & r < r_L \\ 0 & \text{otherwise} \end{cases}$$

with A being large compared to $1/\delta t$. But again, from the numerical point of view this

has the problem that the force is exactly zero at $r_L + \varepsilon$ and nonzero and large at $r_L - \varepsilon$ which causes numerical errors and irregular behavior. Also from the physics perspective, this potential is bad because the direction of the force is ambiguous at the origin. A physically meaningful force would either become infinite so that the other particle could never reach $r = 0$ or it would go to zero so that the force direction wouldn't matter at $r = 0$

$$\text{Physical: } \lim_{r\to 0} F(r) = 0 \quad \text{or} \quad \infty$$

Since $F(r) \to \infty$ is again numerically problematic, the only choice is $F(0) = 0$. To get $F(0) = 0$ we may choose the potential to be an inverted parabola like

$$V \sim \begin{cases} A(r - r_L)^2 & r < r_L \\ 0 & \text{otherwise} \end{cases}$$

Of course we can also take $V \propto 1/r^n$ and regularize it to something like $1/(r^n + \varepsilon)$, but the resulting force is long-range, because it falls like a power-law. We may again cut this force off beyond r_L, but all these choices make the repulsive force harder to control and make assuring that the links stay outside of $2r_L$ distance harder to implement. Also, to fix the numerical force discontinuity issue when $r \to r_L$ we have to have a potential which is smooth at r_L. Thus instead of the patchwork encountered above, the numerically sane alternative that provides both a controlled, localized force and with an adjustable amplitude without having to patch different functions together and without the need for regularization is the Gaussian potential. The Gaussian provides a force which is smooth everywhere, including at the origin and is therefore an ideal choice for numerical simulations. The thickness of the links will then be twice the width of the Gaussian potential and the amplitude is the strength of the repulsion.

1. Gaussian-type Repulsive Force

We argued above that one of the most appropriate choices of repulsive potential which is tailored for our purpose is the Gaussian potential. We will assume that every line segment in one link has a repulsive Gaussian interaction with every segment of another link. Consider a

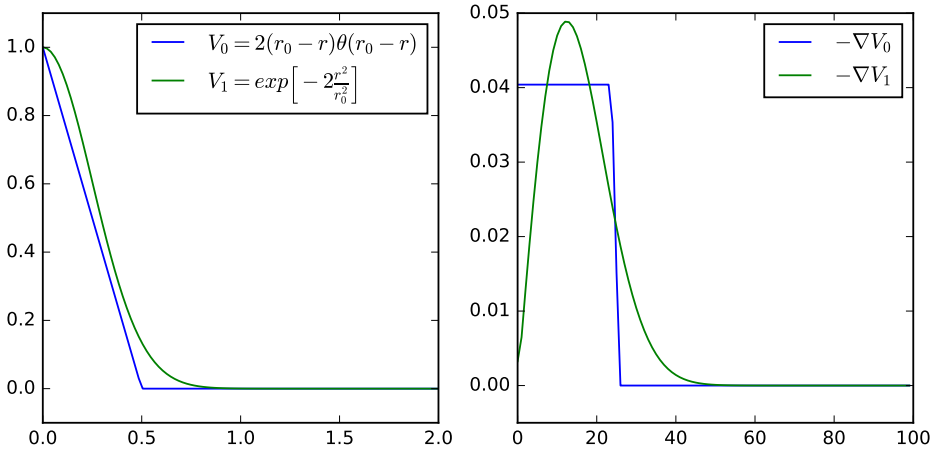

FIG. SI.12. Comparison of the linear incline and the Gaussian potential. While looking similar in shape, the force from the Gaussian is much more well-behaved at $r = r_L$ and $r = 0$ and is better suited for numerical analysis.

segment of link a. Denote the value of the length parameter parametrizing the path γ_l of the link by s_l. The tangent vector to this infinitesimal segment $d\vec{x}_l(s_l)$ interacts with segments of other links via

$$dV_{lm} = ds_l ds_m A \exp\left[-\frac{|\vec{x}_l - \vec{x}_m|^2}{4r_L^2}\right]. \tag{SI.13}$$

Here A is the amplitude of the force and r_L is the cross-sectional radius of the links. Thus, the total interaction between segments of a and b is found by integrating over the paths of the two links

$$V_{lm}[\gamma_l, \gamma_m] = \int_{\gamma_l}\int_{\gamma_m} dV_{lm} \tag{SI.14}$$

which depends on the two paths γ_l and γ_m. The full link-link repulsion potential is then

$$V_{LL} = \sum_{l \neq m} V_{lm} \tag{SI.15}$$

Other potentials of the form $\exp[-|\vec{X}_i - \vec{X}_j|^n/(2r)^n]$ with $n > 2$ may also be used. The higher the n, the steeper will be the potential.

D. Node Repulsion

To add node dynamics to the system, we will define the node positions $\vec{X}_i$ as dynamical variables. The full node-node repulsion potential is

$$V_{NN} = \sum_{i \neq j} V_{ij}, \qquad V_{ij} = A_N \exp\left[-\frac{|\vec{X}_i - \vec{X}_j|^2}{4r_N^2} \right]. \tag{SI.16}$$

where i, j are node indices, A_N is the amplitude of the repulsive potential and r_N is the node radius.

The force from the links on the nodes does not come from a new potential. It is, in fact, contained inside the elastic potential of the links, as the endpoints of the links are attached to the nodes. The only point is that we will allow nodes to be dynamic, instead of fixed. We denoted the tangent vector along link a by $\vec{v}_l(l) = d\vec{x}_l(s_l)/ds_l$. To couple the links with their corresponding nodes we simply need to add a minimal coupling interaction between the end-point tangent vectors $\vec{v}_l(0)$ and $\vec{v}_l(s_l)$ with the nodes they attach to. Here s_l denotes the final value of the affine length parameter s_l for link a. The potential for this interaction looks like[2]

$$V_{NL} = c \sum_{i=1}^{N} \vec{X}_i \cdot \sum_{a \in <i>} \vec{v}_l \left(l_l^{(\text{end})} \right) \tag{SI.17}$$

where $< i >$ denotes the set of indices of links connected to node i and $l_l^{(\text{end})}$ is either 0 or s_l depending on which end of link a is attached to node I. We will assume $c = 1$. We will also assume that the mass and friction constant for the nodes is such that the λ_N in the Langevin equation is of the same order as the λ for the link segments in (3).

[2] In general each term could have a different coupling coefficient, but that degree of complexity is unnecessary here.

E. Full Potential Energy

In summary, the total potential energy of the system becomes

$$\begin{aligned} V &= V_{el} + V_{NL} + V_{NN} + V_{LL} \\ &= \frac{k}{2}\sum_l \int ds_l \left|\frac{d\vec{x}_l}{ds_l}\right|^2 + k\sum_{i=1}^{N}\sum_{l\in<i>} \vec{X}_i \cdot \frac{d\vec{x}_l}{ds_l}\bigg|_{s_l=s_l^{(\mathrm{end})}} \\ &+ A_N \sum_{i\neq j} \exp\left[-\frac{|\vec{X}_i - \vec{X}_j|^2}{4r_N^2}\right] + A_L \sum_{l\neq m} \iint ds_l ds_m \exp\left[-\frac{|\vec{x}_l - \vec{x}_m|^2}{4r_L^2}\right], \end{aligned} \quad \text{(SI.18)}$$

where V_{el} is the total elastic potential of all links $l = 1, ..., L$. Each link is modeled as an elastic cylinder with radius r_L, experiencing both internal elastic forces and short-range external repulsive force from other links and nodes. V_{NL} captures the node-link interactions at link endpoints; the non-crossing condition is ensured by a short-range repulsive force in V_{NN} (node-node interaction) and V_{LL} (link-link interaction) modeled as short-range Gaussian potentials whose strength is set by A_N and A_L. In (SI.18) s_l is the length parameter of link l and $\vec{x}_l(s_l, t)$ represents the position of a point along the center of the link at time t; $s_l^{(\mathrm{end})}$ is the parameter value at the endpoints of link l; $l \in < i >$ represents the set of links connected to node i; $\vec{X}_i(t)$ is the position of node i; r_N is the range of the node-node repulsive force; k is the elastic constant of the links.

F. Dynamics and Minimization of the Potential Energy

The total potential energy for the links, which is a function of the set of paths $\{\gamma\}$ that they take, is given by

$$V[\{\gamma\}] = V_{el} + V_{LL} = \sum_l E_l + \sum_{lm} V_{lm}. \quad \text{(SI.19)}$$

To find the ground state of this system, i.e. optimal static configuration, we can introduce strongly dissipative dynamics in the system. As we show below, this will make the equations of motion effectively first order, gradient descent equations.

Consider a point $\vec{x}_l(s_l(t))$ on link a where the length parameter has value $s_l(t)$ and at time t. We can introduce a kinetic energy term to the energy which is the sum of the kinetic

energies of all infinitesimal segments. A segment of length ds_l will have a mass λds_l, λ being the linear mass density, and thus its kinetic energy is

$$dK_l = \lambda ds_l \left| \frac{d^2\vec{x}_l}{dt} \right|^2 .$$

Summing over all segments, we get $K_l[\gamma_l] = \int_{\gamma_l} dK_l$. Friction terms for an infinitesimal segment will have a similar structure. If we assume we are in a viscous medium and that the amount of friction is simply proportional to the length, in the equation of motion for a segment of length dl, kinetic friction f_k appears as

$$df_k(l,t) = \eta ds_l \frac{d\vec{x}_l(l,t)}{dt}.$$

Since the potential depends on both time t and the length parameters s_l, the equations of motion are more complicated than the usual particle equations of motion. We can think of the position vectors $\vec{x}_l(s_l,t)$ as three component (in 3D) fields whose parameters are s_l, t. The equations of motion are derived the same way as for fields. We have a potential energy density that is a function of both the field $\vec{x}_l$ and its derivative with respect to s_l $\vec{v}_l \equiv d\vec{x}_l/ds_l$. The equation of motion therefore includes partial integration with respect to s_l. In summary, if we denote components of $\vec{x}_l$ by x_l^μ, we have

$$\text{EoM}: \quad \lambda \frac{d^2x_l^\mu}{dt^2} + \eta \frac{dx_l^\mu}{dt} = -\frac{\partial V}{\partial x_l^\mu} + \frac{d}{ds_l}\frac{\partial V}{\partial v_l^\mu}. \tag{SI.20}$$

The first term on the right is the familiar $\nabla_{x_l} V$ and will be nonzero only for the repulsive potential V_{lm}. But the second term acts on the internal elastic potential and yields

$$\frac{d}{dl}\frac{\partial V}{\partial v_l^\mu} = k\frac{d^2x_l^\mu}{ds_l^2}.$$

This is measuring the change in lengths of the segments from one point to another. The reason is that when a rubber band is stretched uniformly, the internal points will not move relative to each other because all forces in all pieces are the same and they cancel out. There will be a non-zero force, and hence movement along the rubber band, only if there is a change in the segment tangent vectors $\vec{v}_l = d\vec{x}_l/ds_l$ in adjacent segments, as they measure length

vectors along the band.

We can now derive the Euler-Lagrange equations for a segment of length ds_l around this point using the action $S = K - V$. To do so, we need the variations with respect to both the position of the points $\vec{x}_l$ and the tangent vector $\vec{v}_l \equiv d\vec{x}_l/ds_l(t)$. They read

$$\begin{aligned}\frac{\delta S[\{\gamma\}]}{\delta \vec{x}_l}: \quad \frac{d^2\vec{x}_l}{dt^2} =& \vec{\nabla}_{x_l} V[\{\gamma\}] - \frac{d}{dl}\vec{\nabla}_{v_l} V[\{\gamma\}] \\ =& -k\frac{d^2\vec{x}_l}{ds_l^2} + \frac{A}{2r_L^2}\int_{\gamma_m} ds_m(\vec{x}_l - \vec{x}_m)\exp\left[-\frac{|\vec{x}_l - \vec{x}_m|^2}{4r_L^2}\right] \end{aligned} \tag{SI.21}$$

To allow the system to quickly converge to a local minimum of the potential, we will assume strong friction, meaning that, if the observation time is τ, we will assume $\lambda\tau \gg 1$ and will discard the acceleration term. So in the end the equation whose dynamics we will simulate to find local minima of this system is

$$\eta\frac{d\vec{x}_l}{dt} = -\vec{\nabla}_{v_l} V + \frac{d}{ds_l}\vec{\nabla}_{v_l} V. \tag{SI.22}$$

Once the dynamics stops, all forces vanish and we will end in some local minimum.

G. Curse of Local Minima

Notice that the equations above stop the dynamics at any local minimum. Similar to many other coupled systems, such as spin glasses [23, 24], the energy landscape of this system is rough and consists of exponentially many locally optimum states which may themselves lie far from the globally optimum configuration. Therefore, we will use Simulated Annealing to allow the system to escape local minima and find close-by, potentially lower local minima.

The amplitude of repulsion among links, as well as the choice of number of segments play a role in how well the system can manage to get out of local minima. Choosing a high number of link segments to mimic the continuity of links, will make it harder for links to pass through each other. Subsequently, tunneling between local minima gets more difficult and the layout will almost always get stuck in a local minimum, where part of the lattice is twisted, or has a kink.

1. Simulated Annealing and Vast Local Minima

The trick is to add some thermal noise to the system. In our case, we can think of the thermal noise literally as thermal fluctuations of the links. We start by assuming that the positions of link segments are only known as a statistical average and that they fluctuate. Thus we replace

$$ds_l(t) \to \langle ds_l(t) \rangle_{T(t)}$$

where $T(t)$ is the temperature of the system at time t. The amount of fluctuation will be proportional to the temperature and the energy of the path. Thus, we have to sum over all paths that the links take and we have to weight each path by the Boltzmann factor based the energy of the path. If we discard the kinetic part of the energy because of high friction, the expression for the Boltzmann weight becomes similar to the "world-line formalism" of polymers. Denote the sum over all possible paths γ_l for link a by $\int [d\gamma_l]$. The partition function for this system, assuming strong friction, is given by

$$Z[\beta] = \int \prod_l [d\gamma_l] \exp\left[-\beta V[\{\gamma\}]\right], \tag{SI.23}$$

with $\beta = 1/T$. This is closely related to the world-line formalism of polymer partition function [2]. Now, instead of the original Langevin equations, the expected values of variables satisfy these equations

$$0 = \left\langle \lambda \frac{d\vec{v}_l}{dt} + \vec{\nabla}_{v_l} V \right\rangle = \int \prod_l [d\gamma_l] e^{-\beta V[\{\gamma\}]} \left(\lambda \frac{d\vec{v}_l[\gamma_l]}{dt} + \vec{\nabla}_{v_l} V[\{\gamma\}] \right). \tag{SI.24}$$

Note, however, that this equation only makes sense if we are close to equilibrium because we are using an equilibrium Gibbs distribution. That is the thermal fluctuations must happen at a time-scale that is much faster than the time-scale of the dynamics so that the thermodynamic averaging makes sense. When this is not the case, the problem still makes sense as a non-equilibrium system, but the expected value equation above will deviate significantly from zero if the temperature is not extremely small. In other words, even when

the dynamics takes the system to the vicinity of a local minimum we have

$$\langle F(t)\rangle_{T,dt} \approx F \pm \delta t \frac{|\partial_t E|}{\sqrt{T}} \sigma(F)_{T,dt}, \tag{SI.25}$$

where F is the value without thermal fluctuations (i.e. at $T = 0$), and $|\partial_t E|$ is the rate at which the system's energy is changing. This means that when the system is not at equilibrium the thermal average taken over a period dt will fluctuate in accordance with the rate of energy change. Since the global minimum cannot be calculated analytically in general, we don't have to worry about near equilibrium description and running the annealing schedule should yield good enough configurations because the fluctuations in (SI.25) allow the system to jump over barriers in the potential energy that are smaller than the σ_F term in (SI.25).

The setup described above is mathematically similar to that of interacting elastic polymers, or more generally fluctuating manifolds in [3] with imaginary noise.

2. Annealing Schedule

We start cooling the system with an exponential cooling schedule and once it "freezes" as $T \to 0$, we will examine the energy. Simulated annealing is not guaranteed to find the global optimum, but it generally finds states that are much better than a local minimum attained by a fixed initial condition[3]. This is related to the fact that the thermal fluctuations may be seen as renormalization of the free energy (similar to the discussion of chapter XI of [19]) which gets rid of small local dents in the potential energy landscape.

Thus, all we need to do is add thermal noise to the system during the dynamics and gradually decrease the level of the noise, hence decreasing the temperature and "anneal" the system. We choose the schedule of the annealing process to be exponential in time

$$T(t) = T_0 \exp[-t/\tau_{\text{cool}}], \tag{SI.26}$$

where τ_{cool} is the cooling exponent, chosen based on the speed of the dynamics. The system

[3] The reason is that the thermal fluctuations allow the system to explore large areas of the energy landscape and the gradient descent introduced as the dynamics will help the system fall into the lowest minimum within that large phase space area. The fluctuations, which allow the system to overcome small barriers between local minima, are effectively smoothening the energy landscape in the sense of (SI.25).

will need the thermal noise for short period of time. T_0 is also chosen based on the parameters of the system. In particular, initially $T_0 > A$ where A is the amplitude in Eq. (SI.19) of the repulsive potential V.

SI 4. CONTRACTIBLE LINKS WITH SELF-REPULSION

Note the $l \neq m$ in (SI.18) in the link repulsion V_{LL}, excluding self-repulsion in links. The reason is that V_{LL} models every link segment ds_l as a soft sphere repelling other link segments in its vicinity. Thus, allowing self-repulsion would interfere with the contraction of elastic links and result in wiring length exceeding minimal possible length. While in most situations avoiding such self-repulsion is desired, in some cases where space is very limited links are pressed against themselves, bending sharply onto themselves as a consequence, resulting in unnatural layouts (Fig. SI.13C). For these situations, we propose a more sophisticated potential energy which models the link segments ds_l as ellipsoids. These ellipsoids will have radii r_L in directions perpendicular to the link segment and radius $|ds_l|/2$ along the link. V_{LL} will then describe the interaction of pairs of ellipsoids, which can be achieved by replacing the simple Euclidean norm $|x_l - x_m|$ by a norm calculated using a "double-ellipsoidal metric" $\sqrt{\Delta x_{lm}^T g^{(lm)} \Delta x_{lm}}$, where $\Delta x_{lm} = x_l - x_m$ (SI 4). For a pair of link segments ds_l and ds_m this metric is given by

$$g^{(lm)} = \left[\left(r^{(l)} {R^{(l)}}^T + r^{(m)} {R^{(m)}}^T \right) \left(R^{(l)} {r^{(l)}}^T + R^{(m)} {r^{(m)}}^T \right) \right]^{-1}, \tag{SI.27}$$

where $r^{(l)} = \left(\frac{ds_l}{2} \middle| r_1 \middle| r_2 \right)$ is the "structure matrix" of ellipsoid l, which is the set of three vectors defining the axes of the ellipsoid (one along the link segment $dsl/2$, and r_1, r_2 perpendicular to ds_l) and $R^{(l)}$ is the rotation matrix transforming these three vectors into the principal x, y, z directions in space. The interaction of ellipsoids has been discussed extensively in the context of polymers and the Gay-Berne potential [4, 25, 26]. It has also been used to describe DNA and other macromolecules as stack of ellipsoids [27–29]. The only difference in our case is that the ellipsoids will stretch and shrink with the links. Using this more accurate double-ellipsoid V_{LL} and allowing $l = m$, resolves the anomalies in the case where space is limited (Fig. SI.13). Finally, note that the outcome of the ellipsoidal potential

is the same as the spherical case when space is not limited as both models are minimizing wiring length, while avoiding crossings, and both have short-range repulsive forces. Thus we will use the spherically symmetric potential in (SI.18) in most cases, unless space is too limited and the layout becomes unnatural.

A. Resolving Self-crossing Using a Gay-Berne Potential

When defining repulsive forces for elastic links, a dilemma. Consider a simple short-range repulsive force, such as hard-core repulsion or smooth versions of it. They all define a minimum distance that any given point on a link would have from all other link points. Here lies the dilemma: We want the links to not have a predefined minimum length –the length is determined by the optimization process–, but on the other hand, we don't want a link to cross itself, either. If we introduce self-repulsion through a fixed radius hard-core repulsion, that radius times the number of discrete link segments determines the minimum link length and it will prevent us from finding optimal wiring length. The resolution that we propose is a repulsive force that adjusts itself to the length of the link.

In directions perpendicular to the link segment, we want the range of the repulsive force to be the link thickness r_L. But along the link segment with length ds, we want the force to only reach a distance $ds/2$. This assures that the segment does not push its adjacent segments away and, thus, doesn't interfere with the link shrinking process, which is essential in the wiring length optimization. To do so requires defining a much more complex type of force than spherical repulsion. We need to define a force based on the interaction of two ellipsoidal force fields. We call this the "double-ellipsoidal" basis, and it will require defining a metric for every pair of interacting link segments to measure how close they are relative to the link orientation, segment length and the link thickness. This type of problem has been encountered before in polymer dynamics and the modified potentials are known as Gay-berne potentials [4, 25, 27].

B. Double-Ellipsoidal Metric

The current form of the repulsive force between two links i, j is a function of the distance between points x_i and x_j along the links. Defining a distance requires defining a metric.

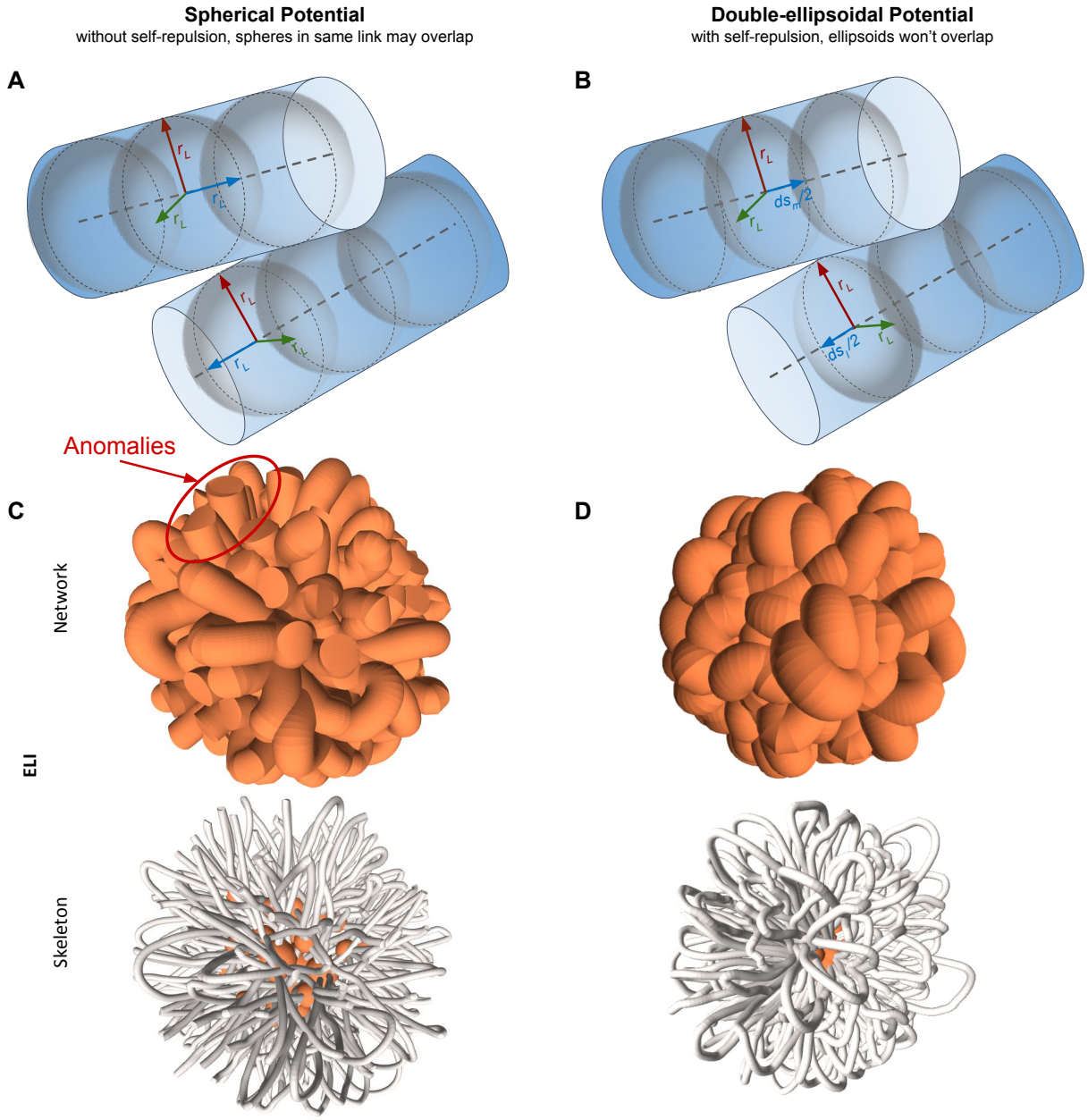

FIG. SI.13. **Spherical vs Double-Ellipsoidal Potential:** If the forces between links l, m depend on the Euclidean distance between link segments, it is as if each link segment generates a spherically symmetric potential around it and the overlap of these potential fields defines the interactions **(A)**. We must exclude self-repulsion within links in this case if we want the links to shrink to their minimal lengths. Excluding self-repulsion is not an issue as long as no link is pushed against itself. However, it poses a problem when space is very limited, for instance when nodes are fixed and links are very thick **(C)**. Lack of self-repulsion removes the penalty for links bending into themselves, resulting in sharp cusps and unnatural layouts. We resolve this issue by using a more sophisticated potential field in link segments, and including self-repulsion. If instead of a spherical potential field, each segments emanates a field which stretches and shrinks together with the link, we can make sure that the link segments of the same link exclude each other, while at the same time allowing the link to shrink to its minimal length. thus we define an ellipsoidal field stretching half the segment length $ds_l/2$ along the link l and r_L in the cross-section of the link **(B)**. The potential energy, then, arises from overlap of two ellipsoidal potential fields of a pair of links l and m. **(D)** shows the same network as in **(C)** but this time using the double-ellipsoidal potential. Notice the smooth nature of the layout in **(D)** and the absence of the sharp ends resulting from self-crossing in **(C)**.

Naively, one might expect that this metric is the Euclidean space metric, but note that the strength of the force depends on the scale defined in this metric. For instance, in the Gaussian potential for links with thickness r_L the width of the Gaussian can be thought of as part of the metric function $g(X, Y)$. In other words, the metric yields

$$g(\Delta x, \Delta x) = \Delta x^i \Delta x^j \frac{\delta_{ij}}{(2r_L)^2}.$$

The denominator comes from the thickness of the two links. When the thicknesses are different, we must have

$$g(\Delta x, \Delta x) = \Delta x^i \Delta x^j \frac{\delta_{ij}}{(r^{(i)} + r^{(j)})^2}.$$

To generalize this to the case of ellipsoidal potential, we must first understand how this metric relates to projections along different directions. The inverse of the metric can be related to these projections. We want a definition of the inverse metric which would yield $\left(r^{(i)} + r^{(j)}\right)^2$ in the spherical case.

The inverse metric is defined on the tangent space as a symmetric rank 2 tensor $g^{-1} = (g^{-1})^{\mu\nu}\, \partial_\mu \partial_\nu$. The components of the inverse metric come from three projections onto three vectors v_i

$$\left(g^{-1}\right)^{\mu\nu} = V_\rho^\mu \delta^{\rho\lambda} V_\lambda^\nu = \sum_\rho V_\rho^\mu V_\rho^\nu. \tag{SI.28}$$

For each pair of links i, j if we define the vectors as $V_\rho^\mu = \left(r^{(i)}{}_\rho^\mu + r^{(j)}{}_\rho^\mu\right)$, $r^{(i)}$ and $r^{(j)}$ are each a set of three mutually orthogonal vectors defining the axes of the ellipsoids. When we have spheres, all three vectors $r_\rho^{(i)}$ and $r_\rho^{(i)}$ have the same magnitude. The spherical case should reduce to the usual metric with the radii $r^{(1)}$ and $r^{(2)}$ of the spheres appearing in the metric along the diagonal, as

$$\left(g^{-1}\right)^{\mu\nu} = |r^{(i)} + r^{(j)}| \sum_\rho \hat{r}_\rho^\mu \hat{r}_\rho^\nu = |r^{(1)} + r^{(2)}| \delta^{\mu\nu}.$$

This is because $(g^{-1})^{\mu\nu}$ is a Hermitian matrix with all eigenvalues equal to $|r^{(1)} + r^{(2)}|$ and is therefore proportional to the identity matrix I. The identity matrix is invariant under all rotations and so, this metric is diagonal in all bases. We will use this as a test for the consistency of our double-ellipsoidal metric below.

C. Gay-Berne Potential Energy

We start by defining the "orientation tensor" of the ellipsoid. This tensor is the set of vectors defining the major axes of the ellipsoid and in the orthonormal basis of the axes of the ellipsoid it becomes diagonal. Let us denote everything in the basis of the ellipsoid by a prime as x', y', z'. The orientation tensor in this basis is

$$r'^{(i)} = \left(r'_{x'} \middle| r'_{y'} \middle| r'_{z'} \right) = R \begin{pmatrix} |r'_{x'}| & & \\ & |r'_{y'}| & \\ & & |r'_{z'}| \end{pmatrix}. \tag{SI.29}$$

The set of orientation vectors $r^{(i)}$ with components ${r^{(i)}}_{\mu}^{\nu}$ is a (1,1) tensor and transforms under rotations R as $r \to R^{-1} r R$. For the spherical case, this doesn't matter, because the orientation tensor $r^{(i)} = |r| I$ and is invariant under rotations. But other cases where the ellipsoid has different radii in different directions require us to find the transformations that take a diagonal $r = \mathrm{diag}(r_x, r_y, r_z)$ from the basis of the ellipsoid to the global x, y, z. In practice, we know the axes of the ellipsoid in the global coordinates $\mu = x, y, z$ as three vectors with components ${r^{(i)}}_{x'}^{\mu}, {r^{(i)}}_{y'}^{\mu}, {r^{(i)}}_{z'}^{\mu}$. However, the other index x', y', z' is defined in the basis of the ellipsoid and we still need to transform that index to the global coordinates. To reiterate, we have the three columns of $r^{(i)}$ as vectors in the global coordinates μ, we already have the row index transformed into the global basis and don't need to transform it again. But we still need to transform the ellipsoid μ' index, which involves a rotation similar to the rotation that would act on $r^{(i)}$ to take the μ index to the local μ', just transposed. How do we efficiently find this transformation? To do this, note that $r^{(i)}$ is the result of acting with one rotation R from the one side on a diagonal matrix with the radii of the ellipsoid $r'_{\mu'}$ on its diagonal, that is

$$r^{(i)} = \left(r'_{x'} \middle| r'_{y'} \middle| r'_{z'} \right) = R \begin{pmatrix} |r'_{x'}| & & \\ & |r'_{y'}| & \\ & & |r'_{z'}| \end{pmatrix}. \tag{SI.30}$$

Thus, if we find the magnitude of each column and divide the column by it, we find R. Then we simply multiply $r^{(i)}$ by the transpose of R from the right in order to go from the μ' basis to the global μ basis,

$$\tilde{r}^{(i)} = r^{(i)} R^T.$$

$\tilde{r}^{(i)}$ has components $\left(\tilde{r}^{(i)}\right)^{\mu}_{\nu}$ where both indices are in the global coordinates. Note that this step is crucial when dealing with two ellipsoids with different orientations, because, although (SI.28) has an inner product on the local index ρ, the basis in which $r^{(i)}$ and $r^{(j)}$ are diagonal are the orientation bases of each of the ellipsoids they represent. Therefore, both have to be first transformed into a similar basis (e.g. the global basis) before they can be added in V^{μ}_{ρ}. Once we have these transformed orientation vectors, we can compute the inverse metric which we need for the double ellipsoid potential. It is easy to check that for a sphere $\tilde{r}^{(i)}$ is diagonal and thus the inverse metric remains a scaled Euclidean metric. Thus, the final step is to combine the transformed orientation tensors and construct the metric for the pair i, j as

$$g^{(ij)}_{\mu\nu} = \left[\left(r^{(i)} R^{(i)^T} + r^{(j)} R^{(j)^T}\right)\left(R^{(i)} r^{(i)^T} + R^{(j)} r^{(j)^T}\right)\right]^{-1}_{\mu\nu} \tag{SI.31}$$

The potential energy is then defined using the new metric

$$\begin{aligned} \Delta x^{(ij)} &= x^{(i)} - x^{(j)} \\ V_{ij} &= A \exp\left[\left(g^{(ij)}_{\mu\nu} \Delta x^{(ij)^{\mu}} \Delta x^{(ij)^{\nu}}\right)^{p/2}\right] \end{aligned} \tag{SI.32}$$

where sum over μ, ν is understood (i.e Einstein notation).

D. Implementation of the Double-Ellipsoid Potential

To construct the ellipsoids, we first need to find two vectors perpendicular to the dl vector going along the link segment. This can be done using the cross product. For example, the two vectors can be found through

$$\hat{r}_1 = \hat{x} \times \hat{dl}, \quad \hat{r}_2 = \hat{r}_1 \times \hat{dl}, \tag{SI.33}$$

and then define $r_1 = r_L\hat{r}_1$ and $r_2 = r_L\hat{r}_2$. Finally, to ensure that the cross product in $\hat{r}_1$ doesn't vanish, instead of $\hat{x}$ we should choose the direction in which dl has the smallest projection. In terms of time complexity, brute-force matrix inversion is $O(n^3)$ and the dot product using the metric is $O(n^2)$, while the spherical case had a simple dot product for $|\Delta x|$, which is $O(n)$. Here $n = 3$ is the dimensions of the vectors. However, the main part of the time complexity comes from iterating over N_s different link segment points and forming interaction pairs. Since that part is not different between using the spherical versus Ellipsoidal forces, the overall complexity only changes by a factor, and the exponent of N_s will not change.

SI 5. PHASE SPACE AND ORDER PARAMETERS

A. Curvature

The curvature of a curve $\gamma(s)$ measures how much the tangent of the curve $T(s) \equiv \gamma'(s)$ changes along the path. Here s is an affine parameter, like the arc-length. Hence, the extrinsic curvature $\kappa(s)$ is given by the magnitude of the second derivative of the curve

$$\kappa(s) = |T'(s)| = |\gamma''(s)| \tag{SI.34}$$

The path is defined as $\gamma(s) = (x(s), y(s), z(s))$ in Cartesian coordinates and the tangent related to the gradient of $\gamma(s)$,

$$T(s) = \frac{d\gamma(s)}{ds} = \vec{x}'(s) \cdot \vec{\nabla}\gamma(s) = x'(s)\hat{x} + y'(s)\hat{y} + z'(s)\hat{z}. \tag{SI.35}$$

Since s is the affine parameter and $ds^2 = dx^2 + dy^2 + dz^2$, we always have $|T(s)| = 1$. For arbitrary parametrization, we have $T = \gamma'/|\gamma'|$. Therefore, just as in circular motion, $T'(s)$ will be perpendicular to $T(s)$ and related to the radius of curvature. Taking the second derivative, the curvature can be summarized as

$$\kappa = \frac{|\gamma' \times \gamma''|}{|\gamma'|^3}. \tag{SI.36}$$

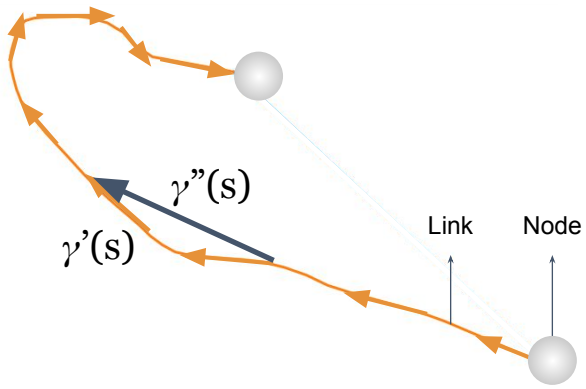

FIG. SI.14. Link curvature, which is equal to the inverse of the radius of curvature, is related to the angle between the tangent vector $\gamma'(s)$ and the change in the tangent vector $\gamma''(s)$. This can be estimated discretely in numerical simulations.

In practice, to measure link curvature, we can measure the cross product of the tangent vector $\gamma'(s)$, calculated for every link segment, and the second derivative $\gamma''(s)$ and divide by $|\gamma'(s)|^3$. Note that $\kappa(s) = 1/R(s)$ with $R(s)$ being the radius of curvature. But, taking the same curve and scaling it up will increase the radius of curvature and decrease the curvature. Therefore, to compare curvature in networks of various sizes we rescale the curvature by the average link length $\langle l \rangle$ and measure $\langle l \rangle \kappa(s) = \langle l \rangle / R(s)$.

1. *Regularizing Singularities in Curvature*

In numerical simulations, the links are broken into segments with finite length. Ideally we want these segments to have comparable lengths and be uniformly distributed over the link. In practice, however, some segments may become tiny due to numerical fluctuations. One tiny piece will lead to a large amount of curvature and corrupt the whole curvature vector. We know that there should be no sharp dents in the simulations. We must mitigate this to avoid artifacts caused by it. We therefore regularize the segment lengths $ds_n \to ds_n + 0.1l/n_s$ by a fraction of the total link length l divided by the number of segments n_s to ensure that tiny pieces do not contribute too much.

B. Relaxation Time

We documented that around the predicted crossover point in FUEL (where both nodes and links are free to move) the relaxation time grows colorblue relatively large. In the weakly

interacting regime node repulsion dominates and determines the layout, while in the strongly interacting regime link repulsion dominates and node repulsion is vanishingly small. Near the transition point, many links are pushing against each other, driving nodes apart from each other and thereby reducing the node repulsion dramatically. At this transition point, the system is switching from a node repulsion dominated phase to a link repulsion dominated one, node repulsion has decreased significantly, while link repulsion is just catching up. Consequently, the relaxation time grows significantly near the crossover. ELI lacks this peak because in ELI nodes are fixed and there is no competition between link and node repulsion.

Note that care must be taken when measuring the relaxation time. Aside from having fixed criteria for when to stop the simulations (e.g. how small the fluctuations in the potential energies must become), we must also make sure that the time units are the same for all simulations. The equations of motion are gradient descent equations and time units will be involved in corrections to positions of nodes and link segments. Therefore, one needs to have an invariant way of defining the time units to determine if there is any significant difference in relaxation times at different thicknesses and values of other parameters. In our simulations, the time step, dt, is determined through a series of functions of other variables. It is defined as

$$dt = c_1(R) \min\{\frac{r_N}{\max F_N}, \frac{r_L}{\max F_L}\},$$

where R is the total edge lengths of the layout and $c_1(R)$ itself is determined from the fluctuations of the layout size and the largest length scale of the problem

$$c_1(R) \equiv \varepsilon_0 + c_0 \frac{\max\{r_N^2, r_L^2\}}{\sigma_R + |\Delta R| + \varepsilon_1},$$

and ε_i are small regularizers. Thus, since the layout size R is proportional to the relevant dominant length scale $r_{\max} = \max\{r_N, r_L\}$, we find that approximately

$$c_1 \propto \frac{r_{\max}^2}{R}, \qquad dt \propto c_1 \times \frac{r_{\max}}{F_{\max}}.$$

Since the repulsive forces and the elastic forces cancel, they will have the same size. The repulsive forces have an exponential piece $F = AF_0 = Ar/r_{\max} \exp[-r^2/r_{\max}^2]$ which is dimensionless, but they also have an amplitude factor A, which we are scaling as $A = A_0 r_{\max}$

in our simulations, and therefore A scales linearly with radius. Since the structure of both node and link repulsion is this way, all the above arguments apply to both the weakly interacting as well as the strongly interacting regime. Thus, in both phases we have

$$dt \propto \frac{r_{\max}^2}{R} \times \frac{r_{\max}}{r_{\max} A_0 F_0} \propto \frac{r_{\max}^2}{R}$$

Thus, to define an invariant time unit, independent of the length scale, we have to divide dt by $r_{\max}^2/R$, obtaining the time we show in the Figure 2D of the paper for the relaxation time.

SI 6. STRESS TENSOR

To understand how ELI and FUEL respond to stress we must analyze various components of their "Cauchy stress tensor" [8, 30]. Consider moving one of a node or a point on a link at position x by a small amount $\delta\vec{x}$. The network will respond with a force, which to lowest order in $\delta\vec{x}$ is

$$\vec{F}(\vec{x}+\delta\vec{x}) \approx \vec{F}(\vec{x}) - \delta\vec{x}\cdot\vec{\nabla}\left(\vec{\nabla}v(\vec{x})\right). \tag{SI.37}$$

Here $v(x)$ is the potential energy density and the full potential energy is $V = \int_{\text{Net}} d^3x v(x)$. The term $-\vec{\nabla}\left(\vec{\nabla}v(\vec{x})\right)$ is the Cauchy stress tensor and has components $T_{\mu\nu}(\vec{x}) \equiv -\partial_\mu\partial_\nu v(\vec{x})$, with μ and ν labeling the (x, y, z) components. Note that $v, F_\mu, T_{\mu\nu}$ are all densities. To get the total potential energy, force and stress we need to integrate over the volume of the space. Recall that the potential energy was given by (SI.18). The derivatives of the elastic force can be evaluated similar to deriving variations for the Euler-Lagrange equations. The first derivative yields

$$\begin{aligned}\partial_\mu\left(\frac{k}{2}\sum_l \int ds_l \left|\frac{d\vec{x}_l}{ds_l}\right|^2\right) &= \frac{k}{2}\sum_l \int ds_l \frac{\delta}{\delta x_l^\mu}\left|\frac{d\vec{x}_l}{ds_l}\right|^2 \\ &= -k\sum_l \int ds_l \frac{d^2x_l^\mu}{ds_l^2}.\end{aligned} \tag{SI.38}$$

A second derivative then yields

$$\begin{aligned}\partial_\nu\partial_\mu\left(\frac{k}{2}\sum_l\int ds_l\left|\frac{d\vec{x}_l}{ds_l}\right|^2\right) &= -k\sum_l\int ds_l\frac{\delta}{\delta x_l^\nu}\frac{d^2x_l^\mu}{ds_l^2}\\ &= -k\delta_{\mu\nu}\sum_l\int ds_l\frac{d^2\delta(x^\mu-x_l^\mu)}{ds_l^2}\\ &= -k\delta_{\mu\nu}\sum_l\int ds_l\Big[\vec{\nabla}\delta(x^\mu-x_l^\mu)\cdot\frac{d^2\vec{x}_l}{ds_l^2}\\ &+\frac{dx_l^{\mu'}}{ds_l}\frac{dx_l^{\nu'}}{ds_l}\partial_{\mu'}\partial_{\nu'}\delta^3(x-x_l)\Big],\end{aligned}\tag{SI.39}$$

where we have summed over μ' and ν'. The first term is related to the curvature of the links at point x and the second term is related to their tangent vectors $v_\mu \equiv \frac{dx_l^{\mu'}}{ds_l}$. In terms of components, we need the Hessian matrix of the potential energy

$$\begin{aligned}T_{\mu\nu}(x) =&\partial_\mu\partial_\nu\left(V_{el}+V_{LL}+V_{NN}+V_{NL}\right)\\ =&-k\delta_{\mu\nu}\sum_l\int ds_l\frac{d^2\delta^3(x-x_l)}{ds_l^2}\\ &+A_L\sum_{l\neq m}\iint ds_l ds_m\delta^3(x-x_l(s_l))\frac{-2r_L^2\delta_{\mu\nu}+x_{lm\mu}x_{lm\nu}}{4r_L^4}\exp\left[-\frac{|x_{lm}|^2}{4r_L^2}\right]\\ &+A_N\sum_{i\neq j}\delta(x-X_i)\frac{-2r_N^2\delta_{\mu\nu}+X_{ij\mu}X_{ij\nu}}{4r_N^4}\exp\left[-\frac{|\vec{X}_{ij}|^2}{4r_N^2}\right],\end{aligned}\tag{SI.40}$$

where $\vec{x}_{lm}\equiv\vec{x}_l-\vec{x}_m$ and $\vec{X}_{ij}\equiv\vec{X}_i-\vec{X}_j$. The first row corresponds to the elastic stress in the links, the second to the link repulsion and the third to node forces[4]. Now we need to estimate the value of the integrals over the links. In the weakly interacting regime, the links touch at isolated points and thus the integral turns into the sum over these touching points. In this phase, the main contribution to total stress $\int d^3xT_{\mu\nu}$ comes from the node term, meaning the last term in (SI.40). In the strongly interacting regime, links push against each other along the whole link. Since the direction and magnitude of the forces is random and without bias in any specific direction, we expect the distance of link segments to be distributed based on the potential energy cost of being at that distance. This means that

[4] Note that the first term contains both the elastic forces and the node-link forces, as they are just the elastic forces at link end-points.

we can treat the potential energy V_{LL} of link-link interaction as a distribution for the link segment distances. Evaluating the integral over $\int ds_l ds_m$ in (SI.40) involves two parts: 1) approximating the integral over directions for $\vec{x}_{lm}$; 2) integrating over the length of the links. The sum over b also becomes an integral over all directions for $\vec{x}_{ab}$ because the link is covered on all sides. Based on observations from the simulations, we will assume that for a random network, every piece of a link is subject to the same distribution of link repulsion in all directions. For links that touch each other $|x_{lm}| \approx \langle r \rangle$. Since in the strongly interacting regime the links fill up the space, each link is approximately covered by other links along its entire length. This means that the sum over adjacent links basically defines an envelope which covers the link. Thus, for the directions of the stress, we can write $\vec{x}_{lm} \approx r\hat{n}$. This yields

$$\begin{aligned} \sum_{b \neq a} \iint ds_l ds_m f(\vec{x}_{ab}) &\approx \int d|s_l| \int r dr \int d\hat{n} f(r\hat{n}) \\ &\approx \langle l \rangle \int r dr \int d\hat{n} f(r\hat{n}), \end{aligned} \tag{SI.41}$$

with $\langle l^2 \rangle$ being the average square length of the links, which in the strongly interacting regime will be a numeric factor times the size of the layout squared. Also, note that

$$\int d\hat{n} n_\mu n_\nu = \pi \delta_{\mu\nu}. \tag{SI.42}$$

This can, for example, be seen by writing $\hat{n}$ in the 2D cross-sectional plane of each point on the link in polar coordinates. In this plane $d\hat{n} = d\theta$ and the x, y components in this plane become

$$\begin{aligned} \int_{\text{bulk}} d\hat{n} n_x n_x &= \int_0^{2\pi} d\theta \cos^2\theta = \pi \\ \int_{\text{bulk}} d\hat{n} n_x n_y &= \int_0^{2\pi} d\theta \sin\theta \cos\theta = 0 \\ \Rightarrow \int_{\text{bulk}} d\hat{n} n_\mu n_\nu &= \pi \delta_{\mu\nu}. \end{aligned} \tag{SI.43}$$

This result is valid in higher dimensions as well, as using spherical and generalized polar coordinates easily show. The result (SI.43) was for the bulk. We will also need to calculate

this at the boundary, where $\hat{n}$ may only be pointing into the network. As near the surface of the network, the normal vector for network elements near the surface can only point into the network, the orientation of $\hat{n}$ cannot take all angles between 0 and 2π, but rather at most half of the angles. Suppose the boundary is the x-axis and the upper half-plane is inside the network. In this case, θ runs between 0 and π instead of 2π and we have

$$\begin{aligned}\int_{\text{bdy}} d\hat{n} n_x n_x &= \int_0^\pi d\theta \cos^2\theta = \frac{\pi}{2} \\ \int_{\text{bdy}} d\hat{n} n_x n_y &= \int_0^\pi d\theta \sin\theta\cos\theta = 0 \\ \Rightarrow \int_{\text{bdy}} d\hat{n} n_\mu n_\nu &= \frac{\pi}{2}\delta_{\mu\nu},\end{aligned} \tag{SI.44}$$

where "bdy" means for links on the surface boundary of the layout.

A. Stress from V_{el}

With regards to the terms from V_{el} in (SI.40), in the strongly interacting regime at every point in the bulk link segments are isotropically distributed. To evaluate the derivatives of the delta function, if we regularize the delta function and write it as a sharp Gaussian, we see that the affine parameter s_l is like r and $|dx_l| = ds_l = dr$. To see this, we start with the substitution

$$\delta^3(x - x_l) = \lim_{b\to\infty}\left(\frac{b}{\pi}\right)^{3/2} \exp[-b|x - x_l|^2].$$

Since the delta function is zero when x and x_l are different, the integral and sum $\sum_l \int ds_l$ in the second line of (SI.40) only needs to be evaluated where $x_l \approx x + \delta x_l$. Although δx_l is infinitesimal, since the width of the regularized delta function is going to zero, we can treat $\sqrt{b}\delta x_l$ as an arbitrary vector which can be integrated to infinity. As a result, in the bulk of the network we get

$$\sum_l \int ds_l \frac{d^2\delta^3(x - x_l)}{ds_l^2} \approx \lim_{b\to\infty}\frac{4b}{\pi^{1/2}}\int_0^\infty dr\partial_r\left(r^2\partial_r \exp[-br^2]\right) = 0, \tag{SI.45}$$

where we have used spherical coordinates and $r \equiv \sqrt{b}|\delta x_l|$. The isotropic link distribution in the bulk therefore results in the same result as $\partial_\mu\partial_\nu V_{LL}$. Similarly, at the boundary, we

still have the same radial part and stress vanishes again. This vanishing of terms is similar to stress in fluids: Inside the fluid all local stress diffuses away in currents and the fluid equilibrates. Of course, we still need to determine the V_{LL} contribution to fully understand if the strongly interacting regime really behaves like a fluid.

B. Stress from V_{LL}

Now we need to determine $\langle r \rangle$ and $\langle r^2 \rangle$ where $\langle r^n \rangle$ is the average of distance between link segments to power n, taking V_{LL} as a distribution. In the strongly interacting regime the cross-section of links is 2D, so the integral of vectors r run in 2D. For all points in the bulk of the network (i.e. everything but the outer surface) we have

$$\begin{aligned}
b &\equiv \left(4r_L^2\right)^{-1} \\
C &= \int d^2 r e^{-br^2} = \frac{\pi}{b} = 4\pi r_L^2 \\
\langle r \rangle &= C^{-1} \int r d^2 r e^{-br^2} = -2\pi C^{-1} \partial_b \sqrt{\frac{\pi}{b}} \\
&= C^{-1} \left(\frac{\pi}{b}\right)^{3/2} = \frac{8\pi^{3/2} r_L^3}{4\pi r_L^2} = 2\sqrt{\pi} r_L \\
\left\langle r^2 \right\rangle &= C^{-1} \int r^2 d^2 r e^{-br^2} = -C^{-1} \partial_b C = 4r_L^2.
\end{aligned} \tag{SI.46}$$

Consequently, inside the bulk of the network, we have

$$\begin{aligned}
\partial_\mu \partial_\nu V_{LL} &= A_L \sum_{l \neq m} \iint ds_l ds_m \delta^3(x - x_l(s_l)) \frac{-2r_L^2 \delta_{\mu\nu} + x_{lm\mu} x_{lm\nu}}{4r_L^4} \exp\left[-\frac{|x_{lm}|^2}{4r_L^2}\right] \\
&\approx A_L \langle l \rangle \int r dr \int_{\text{bulk}} d\hat{n} \frac{-2r_L^2 \delta_{\mu\nu} + r^2 n_\mu n_\nu}{4r_L^4} \exp\left[-\frac{r^2}{4r_L^2}\right] \\
&= A_L \pi \langle l \rangle \delta_{\mu\nu} \int r dr \frac{-4r_L^2 + r^2}{4r_L^4} \exp\left[-\frac{r^2}{4r_L^2}\right] \\
&= A_L \pi \langle l \rangle \delta_{\mu\nu} \frac{-16\pi r_L^4 + 16\pi r_L^4}{4r_L^4} = 0.
\end{aligned} \tag{SI.47}$$

The two terms coming from V_{LL} in (SI.40) cancel each other and the stress vanishes, similar to inside a fluid where no localized stress survives. Of course, if we cut the network and measure the push on a surface placed inside the network, we have produced a boundary and

the result will be different. Similarly, we expect all the excess forces to propagate to the surface of the network. On the surface, we need to use (SI.44) instead of (SI.43) and we have $\int_{\text{bdy}} d\hat{n} = \pi$, not 2π, and so the terms cancel again

$$\begin{aligned}\partial_\mu \partial_\nu V_{LL} &\approx A_L \langle l \rangle \int r dr \int_{\text{bdy}} d\hat{n} \frac{-r_L^2 \delta_{\mu\nu} + r^2 n_\mu n_\nu}{4 r_L^4} \exp\left[-\frac{r^2}{4r_L^2}\right] \\ &= A_L \pi \langle l \rangle \delta_{\mu\nu} \frac{-8\pi r_L^4 + 8\pi r_L^2}{4 r_L^4} = 0 \end{aligned} \tag{SI.48}$$

C. Stress at the Walls Pushing the Network

As we showed above, the contributions from the elastic forces and the repulsive forces both vanish in the bulk, as well as free boundaries of the network. This suggests that the strongly interacting regime may be acting like a fluid, dissipating internal stress. However, to prove that it is actually behaving like a fluid, we have to prove that if a wall pushes against a part of the network, the local stress generated by displacing that part from equilibrium has the form of hydrostatic stress, meaning that we have $T_{\mu\nu} \approx p\delta_{\mu\nu}$ with a pressure p that depends on the force pushing the network. Only in parts where a link or a node is pushed away from its equilibrium position, we will see stress because the forces on it do not vanish any more. The amount of stress will be the change in the force that the displaced link segment experiences. Thus, we have to assume that before the push the forces canceled (i.e. the gradient $\nabla V = 0$) and now the piece is experiencing a force. In other words, the paths s_l of links are varying slightly, and locally, when the push occurs and we want to calculate the energy cost of this variation. Since the forces were zero, the energy cost is second order in derivatives of V.

To calculate the stress at a wall, let us first solve a simple case, where we of have a link with elastic energy pressing against another link via the Gaussian repulsive force. Considering

the forces on only one link segment, the potential energy is of the form

$$\begin{aligned} V_0 &\sim -\frac{k}{2}\int d^3t\,|x_l(s)-x_l(s+t)|^2 + A_L\int d^3x_m \exp[-(x_l(s)-x_m)^2/(4r_L^2)] \\ -\partial_\mu V_0 &\sim k\left(x_l^\mu(s) - \int d^3t x_l^\mu(s+t)\right) \\ &\quad + A_L\int d^3x_m \frac{x_l^\mu - x_m^\mu}{2r_L^2}\exp[-(x_l(s)-x_m)^2/(4r_L^2)]. \end{aligned} \tag{SI.49}$$

At a wall with normal vector $\hat{n}$, for any radial function $f(r)$ we have

$$\int d^3r f(r) r^\mu = c(r)\hat{n}^\mu,$$

which is because $f(r)$ is an even function of r and thus all components perpendicular to $\hat{n}$ cancel. Thus the integrals become

$$\int d^3(\delta s)(x_l^\mu(s) - x_l^\mu(s+t) \sim ds_l \hat{n}^\mu,$$

and

$$\int d^3x_m \frac{x_l^\mu - x_m^\mu}{2r_L^2}\exp[-(x_l(s)-x_m)^2/(4r_L^2)] \sim g(r_L)\hat{n}^\mu.$$

Therefore, regardless of the details of $g(r_L)$ since the gradient at the wall is always proportional to $\hat{n}^\mu$ and the stress tensor will be

$$T_{\mu\nu} = \partial_\mu\partial_\nu V \propto \partial_\mu \int d^3r f(r)\hat{n}_\mu \sim C\delta_{\mu\nu} + B\int d\hat{n}\hat{n}_\mu\hat{n}_\nu \propto \delta_{\mu\nu}.$$

Thus, even though evaluating the exact form of the stress at the walls in the strongly interacting regime might be complicated, it is diagonal and thus hydrostatic in nature. This, together with the vanishing of stress in other parts of the network suggests that the strongly interacting regime behaves like a fluid. Below, we will discuss how this can be validated numerically.

The repulsive terms from V_{NN} in (SI.40) resemble the pressure tensor discussed in [31] used to describe jamming phenomena and stress networks in granular materials [32, 33]. with nodes that feel each other's force being at a distance $|X_{ij}| \approx \langle r\rangle_N$ and nodes farther apart than $\langle r\rangle_N$ basically not contributing to $T_{\mu\nu}$. Here $\langle . \rangle_N$ means we are using the Gaussian

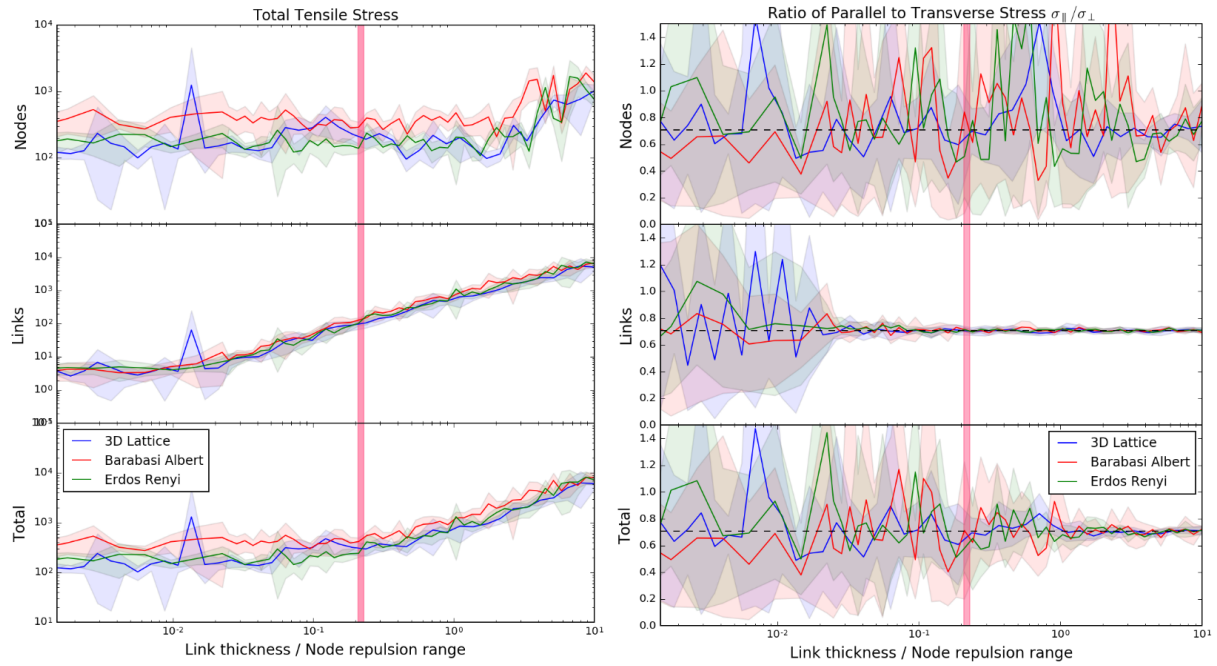

FIG. SI.15. Left: Total stress in an ER and a BA network with 49 nodes and 280 links, and a 3D lattice network with 300 links at various link thicknesses. The solid curve is the average over 10 measurements done by randomly rotating the layout in space and the shaded area shows one standard deviation around this average value. The red strip indicates the approximate transition point. Right: The ratio of parallel and transverse stress in BA, ER and 3D lattice. In the strongly interacting regime, the ratio becomes $1/\sqrt{2}$ (black dashed horizontal line) in all three networks. This indicates transition to hydrostatic stress, meaning that stress is uniformly distributed in all direction like a fluid. The green strip indicates the transition point.

potential V_{NN} as a distribution. Thus, we expect the weakly interacting regime to behave more like an amorphous solid, while, as we found above, we predict that the strongly interacting regime has fluid-like characteristics. Since the links are not constantly moving, it will be more like a gel than a running fluid. This difference in the strong and weak regime should be best visible in the diagonal components of the stress: in the strongly interacting regime we have $T_{\mu\nu} \propto \delta_{\mu\nu}$, i.e. we have hydrostatic stress, so the parallel and transverse components, $\sigma_\parallel$ and $\sigma_\perp$, are proportional to each other. If the stress components σ_x, σ_y and σ_z have similar magnitude, as in hydrostatics, we have $\sigma_\parallel = \frac{1}{L^3}\int_{\text{net}} d^3x|\sigma_y| = p$ and $\sigma_\perp = \frac{1}{L^3}\int_{\text{net}} d^3x\sqrt{\sigma_x^2+\sigma_z^2} = \sqrt{2}p$, consequently $\sigma_\parallel/\sigma_\perp = 1/\sqrt{2}$. If, however, the $\sigma_\parallel/\sigma_\perp$ ratio varies as we rotate the layout, the network possesses solid-like properties.

We examined numerically whether the $\sigma_\parallel/\sigma_\perp$ ratio is constant in the ELI layouts. We find that, as predicted by the theory, $\sigma_\parallel/\sigma_\perp$ fluctuates significantly for node stress at almost all (except at very high) link thicknesses, indicating that it does not satisfy the hydrostatic condition (Fig. SI.15). In contrast, the stress in links shows a remarkably constant $\sigma_\parallel/\sigma_\perp$

ratio starting at thicknesses well below the crossover, indicated by a red strip. At very low thicknesses link stress too shows large fluctuations. In conclusion, the simulations confirm that once link stress dominates over node stress, the ratio $\sigma_{\parallel}/\sigma_{\perp}$ approaches a constant $1/\sqrt{2}$ value and its fluctuations become small. This confirms the fluid or gel-like behavior of the strongly interacting regime.

SI 7. VISUALIZATION PARAMETERS

For a weighted network, there are multiple choices one can make regarding how the weights are related to the link thicknesses. Choosing a linear relation is not always ideal, especially if the distribution of link weights is fat-tailed. For the Flavor Network [11][5], we chose the link thickness to grow with $w^{1/2}$, which means that the cross-sectional area of links is proportional to their weights. This is consistent with what using multiple links would yield. It also makes for a visually more pleasant visual outcome.

A. Relation between Node Radii and Link Thicknesses

To make sure the node size is adequate to make it visible in the presence of its links, the node radius should be related to the link thicknesses. When links cross their node of origin, they cover a portion of its area. When the link thickness is small compared to the node size, the area covered on the link is roughly the cross-sectional area of the link. Thus, for the node to be visible in this situation, its radius R_N must satisfy

$$4\pi R_N^2 > \pi \sum_{l\in<N>} r_l^2 \quad R_N > \frac{1}{2}\sqrt{\sum_{l\in<N>} r_l^2} = \sqrt{k_N} r_{\rm rms} \tag{SI.50}$$

where k_N is the degree of node N. In the extreme case where the link thickness is very large such that one link is covering most of the node area, the cross-sectional area of the link becomes roughly the cross-sectional area of the node itself, meaning $R_N \sim r_l$.

Thus, a safe choice for node size that would satisfy both cases would be $R_N > r_{\rm rms}$. To

[5] The Flavor Network has actually very little variations within each module, as it is a co-occurrence based network. To make the 3D print capability and the visualization more dramatic, we broke this uniformity by randomly shuffling the link weights.

make the nodes prominent in the visualization, we chose

$$R_N = 1.5 r_{\mathrm{rms}}$$

B. 3D Printing

Since the layout of a network will contain links going in all directions, 3D printing them can be occasionally challenging. In particular, 3D printing technologies such as "Fused Deposition Modeling" (FDM) and "Stereolithography" (SLA) which require adding a support structure to the model are possible, but highly impractical. The reason is that a network with randomly oriented links requires a large amount of support structure and removing the supports is rather tedious, if not impossible. We found that the best technology for 3D printing is "Selective laser sintering" (SLS). In this method, at each step a layer of powder is added to the top of the printing bed, and a laser then fuses powder particles in the new layer together and to the previous layer. This way, the powder itself provides a natural support which keeps the model in place and removes the need for support structures entirely. We used SLS through the online service ShapeWays to 3D print networks (main article, Fig. 3).

SI 8. OPTIMIZATIONS

Brute-force calculations quickly become intractable for even medium sized networks. We used a number of strategies to optimize these computations. Many of them involve binning the space in various ways to focus only on the parts where computations are required. Such strategies in some cases reduce the computation time by a factor of 10 or more as they change the time complexity of the problem.

A. Finding Crossings more Efficiently

We only need to check pairs of nodes whose links come close to each other and may thus have a chance of crossing. This means that only nodes $A : (a_1, a_2)$ and $B : (b_1, b_2)$ for which in all three directions x, y, z the four nodes fall within a certain geometric shape. If links were just lines, the sufficient condition for potential crossing would be that A, B form a

tetrahedron with no obtuse angle. Since we only wish to speed up the search for crossings, we can choose a very crude measure. The constraint is that the measure has been evaluated very efficiently and very rapidly, faster than it would take to do the computation for checking the crossing, which involves calculating three cross products and several dot products and square-roots.

B. Partitioning the Space for Efficient Spotting of Potential Link Crossings

We can partition the space into cubes whose edges are longer than a few link thicknesses. Then, most crossings will occur inside these cubes and we don't need to compare link segments which do not cross the same cube. We assign an ID to each link. Then we make a hash-table (for quick look-up) of all the cubes and keep the list of link segments that cross the cube in a list inside this hash-table. The number of the cubes with more than one links crossing them can be very small, depending on how their scale compares to the total space occupied by the network. This partitioning approach allows spotting precisely which segments of which link are crossing. If the network has M, we have to make sure compiling this hash-table and looking up pairs takes less than $O(M^2)$. The hash-table will only have entries for partition cells that had at least one link crossing them. The look-up of hash-table is normally $O(1)$ and $O(k)$ in the worst case, with k being the number of cells occupied by links. Compiling the hash-table is $O(M)$.

C. Time Complexity with and without Partitions

Consider a uniform graph of average degree m, with k cells occupied. We have $M = Nm/2$ links. Each link may stretch several cubes. If they stretch $q \ll k$ regions, we will break the links into q segments. The total number of segments yields the cumulative occupancy of the k regions. Thus each cell has Mq/k links. The total number of checks for crossing is therefore

$$O(M) + O(k) + O\left(k \times \frac{M^2 q^2}{k^2}\right) = O(M) + O(M^2 q^2/k).$$

Assuming that in the worst case the look-up of the hash-table is $O(k)$ (typical case is $O(1)$).

Thus in order for the partitioning approach to be more efficient than the original we must have $q \ll \sqrt{k}$. If our links are straight lines, if the network is embedded more or less symmetrically in all three dimensions, then the maximum number of cubes a link may pass through is $O(k^{1/3})$. For $k \gg 1$ a uniform embedding in d dimensions yields

$$q \sim O(k^{1/d}) \ll O(k^{1/2}), \qquad \text{if: } k \gg 1 \quad \text{and } d > 2.$$

In 3D this method will most likely improve the performance. The amount of improvement is

$$O\left(\frac{M^2 k^{2/3}}{k}\right) = O\left(\frac{M^2}{k^{1/3}}\right).$$

The full complexity depends on how many partitions we choose. Assuming that look-up of the hash-table can be done in linear time $O(k)$ in the worst case, if we choose k such that $1 \ll k \ll M$, we can significantly accelerate finding crossings. Implementing this resulted in a very significant improvement in finding link crossings. For a BA network of 500 links and $m = 2$ minimum degree, the brute force method took about 20 secs, while the regioned approach with maximum 500 spatial regions took only 2 secs. The hash-table is updated every few steps, because the nodes and link segments are moving under the forces. The number of steps before next update to the table is estimated based on the magnitude of forces (the larger the forces, the sooner will be the update).

D. Parallelization using GPU

Once the space is partitioned, the computation of repulsive forces inside each cell can be done independently and may thus be parallelized. Within each cell, we may have a number of link segments and we need to calculate all of their repulsive forces efficiently. We therefore need an architecture that performs matrix computations efficiently. Graphics Processing Units (GPU) are optimized for such computations. High-end GPUs also generally have thousands of compute cores and a single GPU can act as a computer cluster. Rewriting our simulation code to run on a GPU resulted in a 10-50 fold speed-up, on average, on networks with a few hundred links in the strongly interacting regime. To use the GPU, we used Google's machine learning package Tensorflow [34]. The package requires an NVIDIA CUDA

compatible GPU. We ran our simulations on machines with NVIDIA GeForce GTX970m on a laptop, as well as NVIDIA Tesla K80 GPUs on the Google Cloud Platform.

E. ELI and FUEL Computational Complexity

If each link has s segments, the complexity of calculating the elastic forces is $O(sL)$. The brute-force complexity of node repulsion is $O(N^2)$ and link repulsion is $O(s^2L^2)$. Since for a connected network $L \geq N-1$ and $s > 1$, the complexity is mostly due to link repulsion. We can reduce this complexity by realizing that we only need short-range link repulsion to avoid crossings, allowing us to partition the space into cells of size ar_L with $10 > a > 2$ and calculate link crossings in these cells only, as the repulsive forces between links farther than ar_L will be exponentially small. To be specific, if the layout is broken into k cubic cells, each cell contains sL/k link segments, on average, and so the total number of pair interactions has complexity $O(k \times s^2L^2/k^2) = O(s^2L^2/k)$ (SI 8.C). A fixed minimum number of cells, k, is desirable at early stages as links may be very densely packed and a cell of size ar_L may contain too many segments, resulting in memory shortage from storing the $O(s^2L^2)$ interaction matrix in the worst case. When the system has opened up sufficiently, we can break the layout into cells of fixed size ar_L and index which segments of which links cross each cell. The indexing is $O(sL)$ and is updated every few iterations, depending on the magnitude of forces. We only need to calculate forces in cells that contain more than one link. We also shift the cells randomly for each indexing event to remove biases at the cell walls.

In the weak regime using cells speeds up calculations enormously, as the number of crossings is very few and so cells crossed by more than one link are very few. Thus the total complexity will be either $O(N^2)$ from brute-force node repulsion or $O(sL)$ from indexing link segments and elastic forces, depending on which one is bigger. In the strong phase, the links are filling the space. But, as links can shrink and stretch, both with no link self-repulsion (Eq. (1)) and the double-ellipsoid potential (SI C. Eq. (SI.32)) the number of link segments within one cell can vary from 1 segment per cell to $O(s)$ per cell in the worst case. A cell of size ar_L contains a maximum of $O(a^3)$ distinct links. Since links can shrink and stretch, the number of segments of each link that falls in each cell can vary. In the best

case, a cell has 1 segment per link, and in the worst case it contains all s segments. Thus, in the worst case one cell may have $O\left((sa^3)^2\right)$ pairs of interacting link segments. The number n_a of spatial cells with links crossing through them can be estimated from the size of the layout. The cross-section of all the links when the layout has opened up is $R \sim L^{1/2} r_L$ and so dividing by the cell size ar_L, the number of cells becomes $n_a \sim O\left(\left(L^{1/2}/a\right)^3\right)$. Therefore, the worst case time complexity is

$$n_a O\left((sa^3)^2\right) \sim O\left(s^2 a^3 L^{3/2}\right)$$

In practice, because the number of pairs s^2a^6 in a cell may become large, we randomly sample up to a maximum number of pairs $\sim \max(100, s^2a^3)$ in each cell to calculate the repulsive forces and the elastic forces help move the rest of the link segments to the correct positions. This way, for large networks the complexity is mostly due to number of cells $O(L^{3/2}/a^3)$. Our algorithm starts with a minimum number of cells k and switches to cells of fixed size ar_L once the network has opened up enough so that more than k cells are occupied by more than one link.

As in the weak regime link crossings are rare, most cells will contain a single link. Thus, in the weak regime partitioning the space reduces the complexity of link repulsion so much that it may become less than the brute-force complexity of node repulsion, $O(N^2)$. Therefore, with partitioning the time complexity in the weak regime will be $\max\{O(N^2), O(sL)\}$, where $O(sL)$ is the complexity of partitioning the space (same as the elastic forces). In practice, we do not partition the space in every iteration, rather every 50-100 steps.

F. Remarks on scalability

In the strong phase, the reduced complexity $O\left(cL^{3/2}/a^3\right)$ after partitioning can be considerably better than the brute-force $O\left(s^2L^2\right)$. In the weak phase, the full layout is a perturbation to FDL. Thus one can do a fast FDL first –ignoring the internal degrees of freedom of links– and then run the full algorithm to resolve potential link crossings. This can be done for networks with thousands of nodes. With our current simulation code, running on an NVIDIA GPU, a network with a few hundred links takes a few seconds in the weak regime and about ten minutes in the strong phase. A network with 1,000 links in the

strong regime may take one to two hours. The largest network that we have tried so far is a coarse-grained network of the white matter of the mouse brain, with 600 nodes and 20,000 links, which took about a week prior to the recent improvements (Fig.SI.16). As discussed above, while the brute force method would scale as $O(L^2)$ for every iteration, our improved code reduced this to $O(L^{3/2})$ by exploiting the short range of interactions. Examples with the best working settings are provided in the code which is available online.

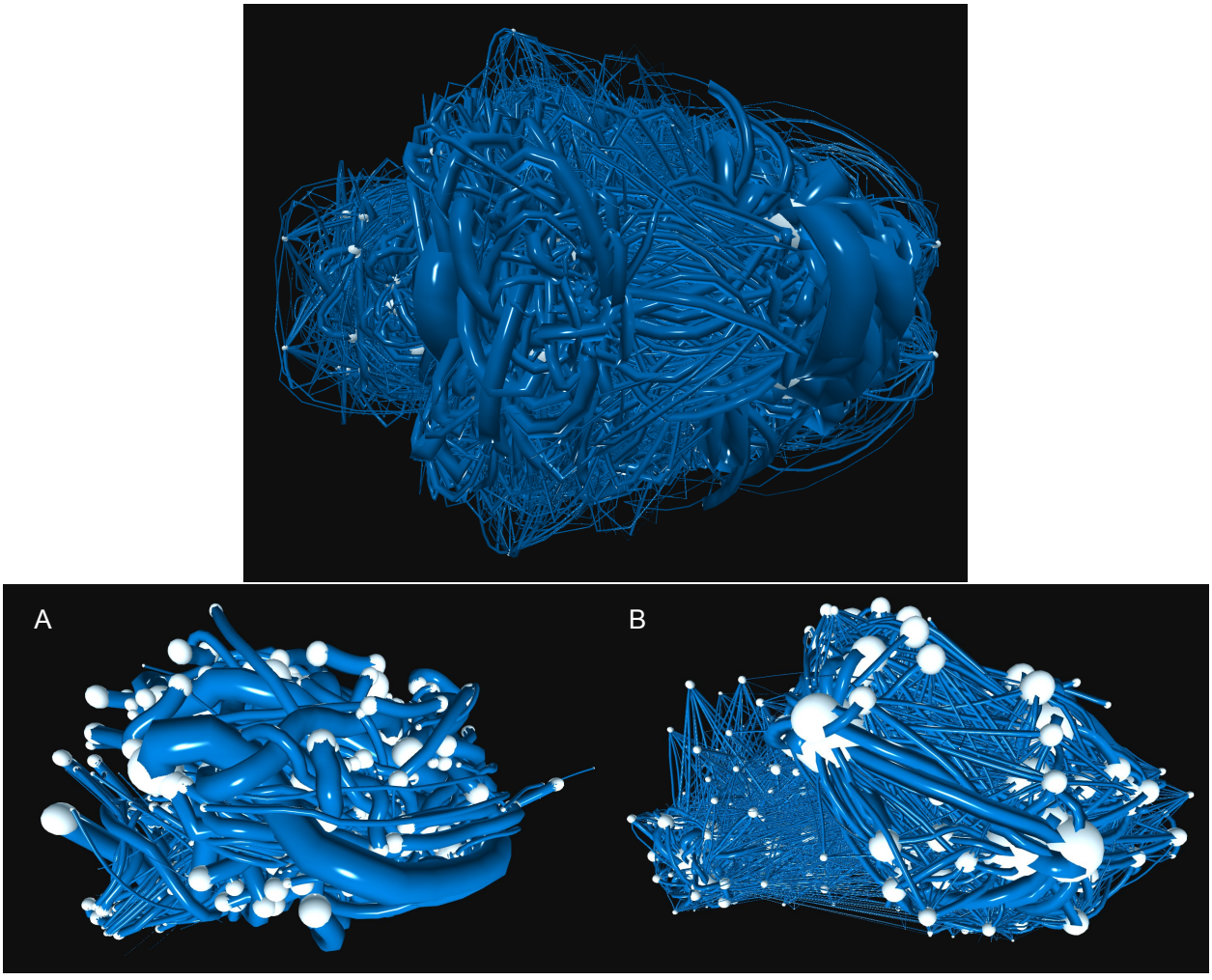

FIG. SI.16. **Simulations based on Mouse Brain White Matter Connectivity:** The Mouse Brain Atlas data, collected by the Allen Institute, shows the trajectories of millions of axon bundles in the white matter of the mouse brain, connecting different regions of the gray matter. We bundle these axon bundles even further to find the total connectivity of large anatomical brain regions, yielding a weighted network of connectivity of brain regions, as well as trajectories of bundles. We then discard the trajectory information, keeping solely the weights (used as link thicknesses) and the location of brain regions (used as nodes). We use this weighted network and run ELI on it, keeping nodes fixed, to generate simulated mouse brains with optimized wiring in 3D. These models serve as null models to compare against the actual trajectories and to answer questions regarding optimality of the white matter connectivity. Depending on the level of detail by which we define anatomical regions in the brain, we obtain different coarse-grained networks of brain connectivity. The top figure shows the top view of a simulation. There are a total 20,000 weighted links, connecting about 700 brain regions at this level of bundling. (A) Side view of the highest level of bundling resulting in 26 large brain regions and 700 weighted links, while (B) one lower level in hierarchy of brian regions, yielding 700 nodes and 20,000 links.

For modular networks or networks with strong hierarchy, one can exploit the hierarchy to lay the network out at multiple scales, thereby reducing the complexity of each run. For non-modular networks, the best heuristic algorithm would be to use k-d tree to partition the space to only resolve crossings within the partitions. We are using a space partitioning with fixed partition sizes in our simulation code as well. Beyond this, for non-modular networks the computation cannot be optimized much further other than large scale parallelization or using the GPU, also implemented in our code and we discuss these recommendations briefly below.

G. Simulation variables and parameters

The strength of the repulsive forces is proportional to the amplitudes A_N and A_L in V_{NN} and V_{LL}, while the strength of the elastic force is proportional to the spring constant k. In order for the model to avoid crossings, the repulsive forces must be much larger than the elastic forces, otherwise the links will be squeezed into each other. We chose $A_N = A_L = A$. Also, for a fixed A and k, a larger network will have proportionately larger elastic forces, as its links are stretched longer. Therefore, to make the ratio of the repulsive forces and elastic forces the same for networks of any size, A must grow with the network size. Thus, for our simulations we choose $A = A_0 \max\{r_N, r_L\}$ and to have a repulsion stronger than the elastic forces we chose $A_0 = 10k$. We checked through simulations that these choices won't affect the outcome of the simulations adversely. We ran many sets of simulations with fixed $A = 100k$ or higher ratios and arrived at the same results. The problem with having a fixed and very large A is that the simulations take a long time to converge. The reason is that the time steps are adaptively calculated from the largest forces in the system to avoid numerical instabilities. We also observed that not scaling A does, indeed, lead to layouts where links are squeezed into each other and the layout does not grow to the correct physical size with links and nodes not crossing each other. Finally, note that the average distance of adjacent nodes and links also depends on the $k/A_{L,N}$ ratios, as we discuss below in sec. SI 11 A 2. This is because the repulsive potential is a sof-core repulsion and the dependence of the "effective" radius of nodes and thickness of links on $k/A_{L,N}$ is non-linear.

SI 9. DIMENSION OF SPACE AND TOPOLOGICAL STATES IN NETWORK LAYOUT

Layouts of networks in 3D can exhibit topological defects, as well as knotted links. In an arbitrary D dimensional space, a node has an exclusion region of the form of a D-dimensional sphere and the cross-section of a link is S_{D-1}, the spherical shell, or boundary of a D-dimensional ball. Intertwined links in 3D, however, can form of knots on the links, because the links remain effectively 1D objects, whereas in $D > 3$ they can always pass by each other. This is because if one considers the segment of links 1 and 2 near the crossing, the segment from link 1 can be embedded in the dimensions x_1, x_2, and the segment from link 2 in x_3, x_4. This makes links 1 and 2 invisible to each other and allows them to resolve the crossing by moving away from the crossing area (which is why knot theory for 1D curves only exists in 3D [12], and in general knotted n-spheres can be untied in $n + 2$ dimensions). So, $D = 3$ is the lowest dimension where any link crossing can be resolved by moving links apart, and also the highest dimension where crossing can lead to knots and topological defects.

In a truly continuous link with infinite hard-core repulsion we cannot resolve the knots, nor is it possible for links to pass through each other to find more favorable configurations. In Eq. (1) we made the repulsive forces smooth, analytical functions with finite maximum repulsive force. This allows nodes and links to tunnel through each other with a suitable amount of thermal noise. However, when the noise amplitude is large (above r_L) and the annealing schedule is such that the noise persists longer than the relaxation time, we do get many twists and knots. In addition to this, we observe "kinks" and "solitons", best seen when laying out 3D lattices (Fig. SI.17 here). The underlying reason is the same as regular solitons: there are multiple global minima for V which are related to each other via discrete symmetries, such as reflection in the x direction. During the dynamics one side of the network may fall into one minimum configuration and the other side into the reflected configuration, resulting in a twist at the boundary of the two layouts. These solitonic solutions are still low energy solutions. Even though they are not a global minimum, their energy is close to the global minimum and their number is more plentiful than the global minima because of the combinatorics. In fact, most layouts of the 3D lattice, including the ones used for Fig.2 turn out to have such twists, but in the log-scale, $\langle l \rangle$ and curvature seem to be similar among

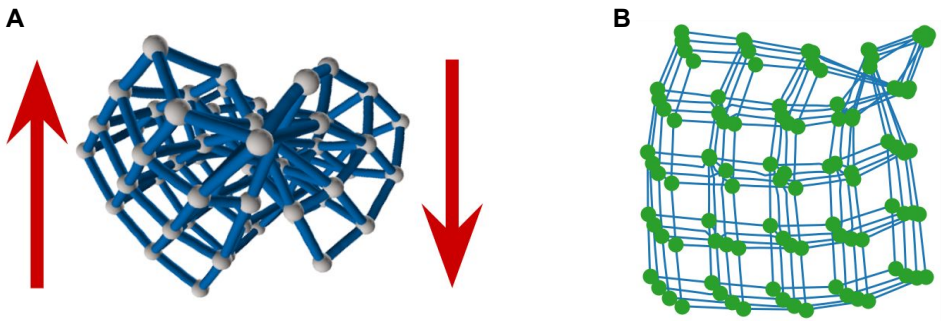

FIG. SI.17. **Solitonic solutions:** A frequent outcome of laying out networks with discrete symmetries, such as 3D lattices due to reflection symmetry, are topological defects: the orientation of two ends of the lattice may be different (A, red arrows pointing to the fact that the orientation of the lattice along the y direction has switched from one end of the layout to the other), resulting in a twist in the middle of the layout. (B) A clearer view of a twist. Twists, like other solitonic solutions, are generally localized, increasing the total link lengths only by a small amount. As a result, these solitonic solutions have energies close to the global minimum. They are encountered frequently during simulated annealing.

them, suggesting that they belong to approximately degenerate energy states.

SI 10. DERIVING THE LOCATION OF THE WEAK-STRONG TRANSITION

We have observed two qualitatively different behaviours in the layout geometry, driven by changes in the ratio r_L/r_N. Here, we analytically derive an equation for the value of $\tilde{r}^c = r_L^c/r_N$, representing the transition point from the weakly interacting to the strongly interacting regime and discern the nature of the transition.

A. Transition from a Node-dominated State to a Link-dominated State

Deep in the weakly interacting regime link crossings are exceedingly rare, hence node repulsion is the dominant force, determining the structure of the layout. Deep in the strongly interacting regime, however, repulsion among links dominates. Indeed, when link thickness, r_L, is larger than the node repulsion range, r_N, in this regime nodes are too far apart to sense each other's repulsion. As the volume of the layout is determined by the exclusion volumes of nodes in the weakly interacting regime and the volume of the links in the strongly interacting regime, our hypothesis is that the change from the weakly interacting to the

strongly interacting regime occurs at a value of $\tilde{r}^c = r_L^c/r_N$ where the node and link exclusion volumes, V_N and V_L, become comparable. To determine this trnasition point, next we will discuss node and link volumes separately.

Total Node Volume: The total node exclusion volume V_N, and thus the effective layout radius R_N, can be found from close-packing of spheres in 3D. The volume occupied by N close-packed spheres is $4\sqrt{2}Nr_N^3$ because spheres can occupy a maximum fraction of $3\pi/\sqrt{2}$ of the volume. We have

$$V_N \equiv \frac{4\pi}{3}R_N^3 = \alpha\pi N r_N^3 \tag{SI.51}$$

where $\alpha = 4\sqrt{2}/\pi \approx 1.8$ assuming the nodes are densely packed in 3D, resulting in

$$R_N = \left(\frac{3\alpha N}{4}\right)^{1/3} r_N = \left(\frac{3\sqrt{2}N}{\pi}\right)^{1/3} r_N \tag{SI.52}$$

Total Link Volume: Finding the exclusion volume of links involves two parts: 1) tubular volume; 2) overlap at end nodes.

First, the node and link volumes become comparable at $\tilde{r}^c < 1$ as each link has a length $R_L \geq 2r_N$ and when $r_L \approx r_N$ the links are filling the layout space. When approaching $\tilde{r}^c$ from the weak phase, links will begin to curve where they are about to intersect other links. However, most links remain approximately straight, the conflicts being resolved by local bendings, until $r_L \approx r_N$. Therefore, we will approximate the volume of each link as a cylinder $v_{0L} = \pi R_L r_L^2$, where R_L is the length of the link. Note that the link length R_L depends on the network connectivity and needs to be calculated separately. We will calculate R_L for scale-free and Erdös-Rényi networks below, after we derive the general condition for the transition.

Links emerging from a point-like node will inevitably overlap up to some radius r_0, which is the minimum radius around the node beyond which its links stop overlapping (Fig. SI.18). For a node of degree k and with link thickness r_L, the region where link overlap is unavoidable is a sphere with radius $r_0(k, r_L)$ around the node. On the surface of this sphere, links are barely touching each other. Let us assume that the surface area of the sphere with radius r_0

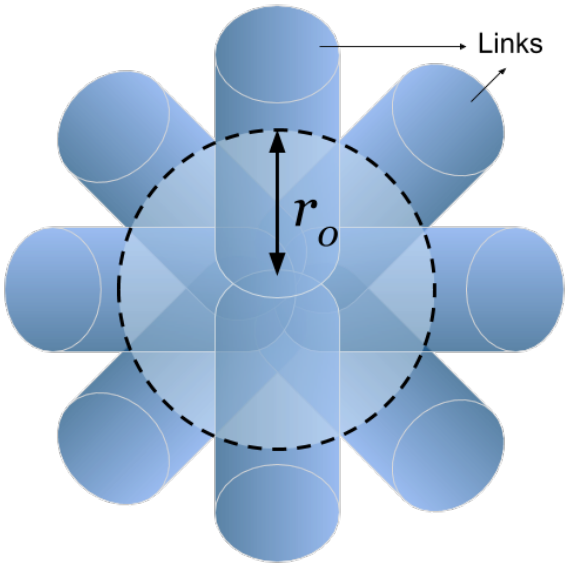

FIG. SI.18. Links emerging from the same node (at the center, node not shown here) will inevitable overlap with each other up to a certain radius, which we denote by r_0. The radius r_0, which is the minimum radius around the node beyond which links emerging from it stop overlapping, depends on the degree k as $r_0 \propto \sqrt{k}$. For any radius smaller than r_0 around the node, the links emerging from it will have some overlap which cannot be resolved.

is completely covered by the total cross-section of the k links. Naively, this means

$$4\pi r_0^2(k, r_L) \approx k\pi r_L^2, \qquad \Rightarrow \qquad r_0 \approx \frac{\sqrt{k}}{2} r_L \tag{SI.53}$$

A slightly more accurate value for $k \gg 1$ can be found by realizing that the linka can arrange themselves in an optimal hexagonal lattice on the surface of the sphere, so the area occupied by each link is $2\sqrt{3}r_L^2$ instead of πr_L^2. Therefore

$$r_0(k, r_L) \approx \sqrt{\frac{\sqrt{3}k}{2\pi}} r_L \approx \frac{\sqrt{k}}{1.9} r_L$$

which is not that different from the approximation (SI.53). Note that the two end-nodes may have different degrees and thus different overlap radii $r_0(k, r_L)$. Thus, to accurately account for the non-overlapping length of links, we will use "stubs": assign $R_L/2$ to each end-node. This way, to calculate the total link volume we sum over the contributions from each node. As for each node we only use R_L length of the link, we avoid double counting. To account

for the link overlap at nodes, the link length used to calculate its cylindrical volume has to be $R_L/2 - r_0$. Having excluded the overlap radius r_0 for each node, we will then add back the link overlap volume $4\pi/3r_0^3$ to find the total link volume. Thus, for node i of degree k_i the total link volume is

$$\begin{aligned} V_{Li}(k) &= \pi r_L^2 \sum_{l \in <i>} \left(\frac{R_{Ll}}{2} - r_0(k, r_L) \right) + \frac{4\pi}{3} r_0^3(k, r_L) \\ &= \frac{\pi}{2} r_L^2 \sum_{l \in <i>} R_{Ll} - \frac{\pi}{3} k^{3/2} r_L^3 \end{aligned} \tag{SI.54}$$

where R_{Ll} is the length of link l, $l \in < i >$ means links l connected to node i, and we used (SI.53). The total volume of the links is found by summing V_{Li} over all nodes i

$$V_L = \pi L r_L^2 \overline{R_L} - \frac{\pi}{3} N \left\langle k^{3/2} \right\rangle r_L^3 \tag{SI.55}$$

where L is the number of links, $\overline{R_L} = \sum_l R_{Ll}/L$ is the average link length, and $\left\langle k^{3/2} \right\rangle = \sum_k p(k) k^{3/2}$ when $p(k)$ is the degree distribution.

B. General Transition Condition

To find the general expression for $\tilde{r}^c = r_L^c/r_N$ we set $V_N = V_L$. In general, $\overline{R_L}$ will be a function of R_N, as we see below. In the random and scale-free networks which we work out explicitly below, we get $\overline{R_L} \approx cR_N$. Using (SI.51), (SI.52) and (SI.55) we have

$$\begin{aligned} V_N \approx V_L, \qquad & \overline{R_L} \approx cR_N \\ \alpha\pi N r_N^3 &\approx \pi r_L^3 \left(c' N^{1/3} L \frac{r_N}{r_L} - \frac{1}{3} N \left\langle k^{3/2} \right\rangle \right) \\ 0 &= \alpha N \left(\frac{r_N}{r_L} \right)^3 - c' N^{1/3} L \frac{r_N}{r_L} + \frac{1}{3} N \left\langle k^{3/2} \right\rangle \end{aligned} \tag{SI.56}$$

where we used (SI.52) to define $c' = (3\alpha/4)^{1/3} c = \left(3\sqrt{2}/\pi\right)^{1/3} c \approx 1.1 \times c$. Rewriting (SI.56) in terms of $\tilde{r}^c \equiv r_L^c/r_N$ yields

$$0 = \alpha \left(\tilde{r}^c\right)^{-3} - c' N^{-2/3} L \left(\tilde{r}^c\right)^{-1} + \frac{1}{3} \left\langle k^{3/2} \right\rangle \tag{SI.57}$$

which represents our main result. Note that Eq. (SI.57) is general, i.e. it is valid for any network topology. We will show below in sec. SI 10 E that (SI.57) has a single positive root satisfying $\tilde{r}^c < 1$, leading to our main result

$$\boxed{\tilde{r}^c = \frac{6A^{\frac{1}{3}}}{A^{\frac{2}{3}} + 12B}, \quad A = -12\left(3\left\langle k^{\frac{3}{2}}\right\rangle + \sqrt{9\left\langle k^{\frac{3}{2}}\right\rangle^2 - 12B^3}\right), \quad B = \left(\frac{3}{4}\right)^{\frac{1}{3}} \frac{cL}{N^{\frac{2}{3}}}} \tag{SI.58}$$

It is important to note that when calculating $\overline{R_L}$ for random or scale-free networks, N and L should correspond to the full network, whereas for lattices and random geometric graphs, N and L need to be replaced by N_u and L_u of the unit cell (the smallest sub-network which is, approximately, repeated in the layout). Indeed, we observe that in lattices with the same structure the transition point is largely independent of the number of nodes in the lattice (Fig. SI.21, bottom).

Finally, note that c depends on the network architecture and layout. We compute approximate values for c for ER and scale-free networks in SI 10 D below.

C. Dependence on degree distribution through $\langle k^{3/2}\rangle$

Here we will explicitly calculate the contribution from $\left\langle k^{3/2}\right\rangle$ in (SI.58). In scale free networks the moments $\langle k^n\rangle$ can be readily calculated from the degree distribution $p(k) = Ck^{-\gamma}$. Using $C = (\gamma-1)k_{\min}^{\gamma-1}$, and $k_{\max} = k_{\min}N^{\frac{1}{\gamma-1}}$ [35] and assuming $n \neq \gamma - 1$ we have

$$\begin{aligned}
\langle k^n\rangle &= C\int_{k_{\min}}^{k_{\max}} k^{n-\gamma} = C\frac{k_{\max}^{n-\gamma+1} - k_{\min}^{n-\gamma+1}}{n-\gamma+1} \\
&= \frac{Ck_{\min}^{n-\gamma+1}}{n-\gamma+1}\left(N^{\frac{n}{\gamma-1}-1} - 1\right) \\
&= \frac{(\gamma-1)k_{\min}^{n}}{n-\gamma+1}\left(N^{\frac{n}{\gamma-1}-1} - 1\right).
\end{aligned} \tag{SI.59}$$

When $\gamma \neq 2$, and $N \gg 1$ we approximately have

$$\begin{aligned}\frac{L}{N} = \langle k \rangle &= k_{\min}\frac{\gamma-1}{\gamma-2}\left(1-N^{\frac{2-\gamma}{\gamma-1}}\right) \\ &\approx k_{\min}\frac{\gamma-1}{|\gamma-2|}\times\begin{cases} N^{\frac{1}{\gamma-1}-1} & \text{if: } 1<\gamma<2 \\ 1 & \text{if: } \gamma>2 \end{cases}\end{aligned} \tag{SI.60}$$

Similarly, for $\langle k^n \rangle$ with $n < 2$ we have

$$\langle k^n \rangle \approx k_{\min}^n \frac{\gamma-1}{|n-\gamma+1|}\times\begin{cases} N^{\frac{n}{\gamma-1}-1} & \text{if: } 1<\gamma<n+1 \\ 1 & \text{if: } \gamma>n+1 \end{cases} \tag{SI.61}$$

We shall now work out some explicit examples.

1. *Scale-free:* $\gamma > 5/2$

From (SI.59) we get

$$\left\langle k^{\frac{3}{2}} \right\rangle \approx \frac{\gamma-1}{\frac{5}{2}-\gamma}k_{\min}^{\frac{3}{2}} = \frac{(2-\gamma)^{\frac{3}{2}}}{\left(\frac{5}{2}-\gamma\right)(\gamma-1)^{\frac{1}{2}}}\left(\frac{L}{N}\right)^{\frac{3}{2}} \tag{SI.62}$$

Therefore, unlike the $1 < \gamma < 2$ case, when $\gamma > \frac{5}{2}$ we have

$$O\left(\left\langle k^{\frac{3}{2}} \right\rangle^2\right) \propto L^3N^{-3} = N^{-1}O\left(B^3\right) \ll O\left(B^3\right) \tag{SI.63}$$

which means that we can ignore $\left\langle k^{\frac{3}{2}} \right\rangle^2$ in A in (SI.58) and write

$$A \approx -i(12B)^{\frac{3}{2}} \qquad\qquad \tilde{r}^c \approx B^{-\frac{1}{2}} \tag{SI.64}$$

and using $c = 3/4$ from sec. SI 10 D 2 we get

$$\boxed{\text{if: } \gamma > 5/2 \qquad \rightarrow \qquad \tilde{r}^c \approx \left(\frac{4}{3}\right)^{\frac{2}{3}} L^{-\frac{1}{2}}N^{\frac{1}{3}}} \tag{SI.65}$$

where $(4/3)^{2/3} \approx 1.21$.

2. Scale-free: $2 < \gamma < \frac{5}{2}$

Again, from (SI.58), to approximate $\tilde{r}^c$, we need A, which requires comparing the magnitude of B^3 and $\left\langle k^{3/2} \right\rangle^2$. In this case from (SI.60) and (SI.59)

$$\begin{aligned} \frac{L}{N} &= \frac{\gamma-1}{\gamma-2} k_{\min} \\ \left\langle k^{3/2} \right\rangle^2 &\propto N^{-2} \left(k_{\min} N^{\frac{1}{\gamma-1}} \right)^3 = \frac{L^{\frac{3}{2}}}{N^2} N^{-3\frac{\gamma-2}{\gamma-1}}. \end{aligned} \tag{SI.66}$$

Since $\gamma > 2$, when $N \gg 1$ we have

$$O\left(\left\langle k^{3/2} \right\rangle^2 \right) = N^{-3\frac{\gamma-2}{\gamma-1}} O\left(B^3\right) \ll O\left(B^3\right) \tag{SI.67}$$

because the exponent $-3\frac{\gamma-2}{\gamma-1} < 0$. Thus, we can ignore the $\left\langle k^{3/2} \right\rangle$, resulting in the same result as $\gamma > \frac{5}{2}$

$$\boxed{\text{if: } 2 < \gamma < 5/2 \qquad \rightarrow \qquad \tilde{r}^c \approx \left(\frac{4}{3}\right)^{\frac{2}{3}} L^{-\frac{1}{2}} N^{\frac{1}{3}}} \tag{SI.68}$$

3. Scale-free: $\gamma = 2$

For the special case of $\gamma = 2$ we have $k_{\max} = k_{\min} N$ and so $L = N \langle k \rangle = k_{\min} N \ln N$. It follows that

$$\begin{aligned} B^3 &= \frac{3}{4} c_{BA}^3 \frac{L^3}{N^2} = \frac{3}{4} c_{BA}^3 k_{\min}^3 N \left(\ln N\right)^3 \\ \left\langle k^{\frac{3}{2}} \right\rangle^2 &= 4 k_{\min}^3 N = 4 \frac{L^3}{N^2 \left(\ln N\right)^3} \\ O\left(B^3\right) &= \left(\ln N\right)^3 O\left(\left\langle k^{\frac{3}{2}} \right\rangle^2 \right) \end{aligned} \tag{SI.69}$$

Since the difference is in $\ln N$, both B and $\langle k\rangle^{3/2}$ may be significant in A in (SI.58), and thus in $\tilde{r}^c$. Thus, we get

$$A = -36\frac{L^{\frac{3}{2}}}{N\left(\ln N\right)^{\frac{3}{2}}}\left(2+\sqrt{4-\left(c_{BA}\ln N\right)^3}\right), \quad B = \left(\frac{3}{4}\right)^{\frac{1}{3}}\frac{c_{BA}L}{N^{\frac{2}{3}}}$$
$$a \equiv \left(2+\sqrt{4-\left(c_{BA}\ln N\right)^3}\right)$$
$$\boxed{\text{if: } \gamma = 2 \qquad \rightarrow \qquad \tilde{r}^c = \frac{6A^{\frac{1}{3}}}{A^{\frac{2}{3}}+12B} = 36^{-\frac{1}{3}}\left(\ln N\right)^{\frac{1}{2}}\frac{6a^{\frac{1}{3}}}{a^{\frac{2}{3}}+c_{BA}}L^{-\frac{1}{2}}N^{\frac{1}{3}}} \tag{SI.70}$$

4. Scale-free: $1 < \gamma < 2$

This case is generally not of much interest for growing networks because its average degree diverges as $N \to \infty$. We have

$$N\left\langle k^{\frac{3}{2}}\right\rangle \approx \frac{\gamma-1}{\frac{5}{2}-\gamma}\left(k_{\min}N^{\frac{1}{\gamma-1}}\right)^{\frac{3}{2}} = \frac{(2-\gamma)^{\frac{3}{2}}}{\left(\frac{5}{2}-\gamma\right)(\gamma-1)^{\frac{1}{2}}}L^{\frac{3}{2}} \tag{SI.71}$$

Defining $c_\gamma = \left(\frac{5}{2}-\gamma\right)^{-1}(\gamma-1)^{-\frac{1}{2}}(2-\gamma)^{\frac{3}{2}}$ and plugging (SI.71) into (SI.58) we find

$$A = -36\left(c_\gamma+\sqrt{c_\gamma^2-c_{BA}^3}\right)\frac{L^{\frac{3}{2}}}{N}, \quad B = \left(\frac{3}{4}\right)^{\frac{1}{3}}\frac{c_{BA}L}{N^{\frac{2}{3}}}$$
$$c_1 \equiv c_\gamma+\sqrt{c_\gamma^2-c_{BA}^3}$$
$$\boxed{\text{if: } 1<\gamma<2 \quad \rightarrow \quad \tilde{r}^c = 6^{\frac{1}{3}}\left(c_1^{\frac{1}{3}}+c_{BA}c_1^{-\frac{1}{3}}\right)L^{-\frac{1}{2}}N^{\frac{1}{3}}} \tag{SI.72}$$

To give an explicit example, using $c_{BA} = 3/4$ for scale-free networks (shown below in sec. SI 10 D 2)

$$\gamma = \frac{3}{2} \qquad \rightarrow \qquad \tilde{r}^c \approx 2.15\cdot L^{-\frac{1}{2}}N^{\frac{1}{3}}. \tag{SI.73}$$

5. Erdös-Rényi Random Graphs

The degree distribution of Erdös-Rényi (ER) random graphs is a Poisson distribution of the form $p(k) = \langle k\rangle^k e^{-\langle k\rangle}/k!$. We may use the moment generating function of Poisson distributions $g(t) = \exp[\langle k\rangle(e^t-1)]$ to calculate $\left\langle k^{\frac{3}{2}}\right\rangle$ using $\langle k^n\rangle = \left.\frac{d^n g(t)}{dt^n}\right|_{t\to 0}$. Since for

$n = 3/2$ we have a fractional derivative, we can use the Laplace transform of the right hand side to compute the moment, which can be further simplified to [36]

$$\begin{aligned} n_0 &\equiv n - \lfloor n \rfloor \\ \langle k^n \rangle &= \frac{1}{\Gamma(-n_0)} \int_0^\infty \frac{\partial_t^{n_0} g(t-s) - \partial_t^{n_0} g(t)}{s^{n_0+1}} ds \Big|_{t\to 0} \\ \left\langle k^{\frac{3}{2}} \right\rangle &= \frac{\langle k \rangle}{\Gamma\left(-\frac{1}{2}\right)} \int_0^\infty \frac{e^{-s} \exp\left[\langle k \rangle \left(e^{-s} - 1\right)\right] - 1}{s^{\frac{3}{2}}} ds \end{aligned} \tag{SI.74}$$

This integral cannot be written in closed form. But we note that when $\langle k \rangle \gg 1$ the Poisson distribution approaches a Gaussian distribution, with its skewness, equal to $\langle k \rangle^{-1/2}$, becoming very small. Therefore, for the purpose of approximating $\left\langle k^{\frac{3}{2}} \right\rangle$ we will replace the degree distribution with

$$\text{if: } \langle k \rangle \gg 1 \quad \rightarrow \quad p(k) \approx \frac{1}{\sqrt{2\pi \langle k \rangle}} \exp\left[-\frac{(k - \langle k \rangle)^2}{2 \langle k \rangle} \right] \tag{SI.75}$$

This yields

$$\begin{aligned} \left\langle k^{\frac{3}{2}} \right\rangle \approx \Gamma\left(\frac{5}{4}\right) \frac{\langle k \rangle e^{-\frac{\langle k \rangle}{4}}}{6\sqrt{\pi}} \Bigg(& 3(\langle k \rangle + 1) I_{-\frac{1}{4}}\left(\frac{\langle k \rangle}{4}\right) \Gamma\left(\frac{3}{4}\right) + 4(\langle k \rangle + 3) I_{\frac{1}{4}}\left(\frac{\langle k \rangle}{4}\right) \Gamma\left(\frac{7}{4}\right) \\ & + 3 \langle k \rangle I_{\frac{3}{4}}\left(\frac{\langle k \rangle}{4}\right) \Gamma\left(\frac{3}{4}\right) + 4 \langle k \rangle I_{\frac{5}{4}}\left(\frac{\langle k \rangle}{4}\right) \Gamma\left(\frac{7}{4}\right) \Bigg) \end{aligned} \tag{SI.76}$$

Where $I_\alpha(x)$ are hyperbolic Bessel functions. For large x they satisfy

$$x \gg 1 \quad \rightarrow \quad I_\alpha(x) \approx \frac{e^x}{\sqrt{2\pi x}} \left(1 - O\left(x^{-1}\right)\right) \approx \frac{e^x}{\sqrt{2\pi x}}$$

Using this asymptotic behavior for $\langle k \rangle \gg 1$ we get

$$\left\langle k^{\frac{3}{2}} \right\rangle \approx \Gamma\left(\frac{5}{4}\right) \Gamma\left(\frac{3}{4}\right) \frac{\sqrt{2 \langle k \rangle}(\langle k \rangle + 1)}{\pi} \approx 0.89 \cdot \langle k \rangle^{\frac{3}{2}} \tag{SI.77}$$

In conclusion, when $\langle k \rangle \gg 1$ in ER networks for A in (SI.58) we have

$$\left\langle k^{\frac{3}{2}} \right\rangle^2 \approx \langle k \rangle^{\frac{3}{2}} \propto \frac{L^3}{N^3} = N^{-1} O\left(B^3\right)$$
$$\Rightarrow A \approx -i(12B)^{\frac{3}{2}} \tag{SI.78}$$

resulting in the same scaling as (SI.64)

$$\boxed{\tilde{r}^c \approx \left(\frac{3}{4}\right)^{-\frac{1}{6}} c_{ER}^{-\frac{1}{2}} L^{-\frac{1}{2}} N^{\frac{1}{3}} \approx 1.46 \cdot L^{-\frac{1}{2}} N^{\frac{1}{3}}} \tag{SI.79}$$

where we used $c_{ER} = 18/35$ calculated in sec. SI 10 D 1. Taken together, we find that independent of network topology (SF for $\gamma > 2$ and ER), we have $\tilde{r}^c \sim L^{-\frac{1}{2}} N^{\frac{1}{3}}$. Using $L \sim \langle k \rangle N$, we obtain that all large networks have $\tilde{r}^c \sim N^{-\frac{1}{6}}$, in other words for large networks the weakly interacting regime disappears.

D. Average Link Length

In order to evaluate the weak to strong transition we need to estimate the average link length $\overline{R_L}$ in (SI.55). Note that we focus on a near optimal layout, based on our model with short-range node repulsion. In some networks, such as geometric graphs and lattices, the optimal layout consists of links that do not stretch to a length comparable to the layout radius R_N, but rather stay within a local region, or a few "unit cells" (the smallest sub-network which is repeated to make the full network) in the case of lattices. In other networks, such as scale-free and random networks, links may extend the whole layout such that the average link length, $\overline{R_L}$, is comparable to the layout radius, R_N.

1. Average Link Length in Erdös-Rényi Random Networks

Consider an Erdös-Rényi (ER) random network with N nodes and L links, where $L > N$. In FDL, and in ELI/FUEL, any node may have a link to any other node in the layout. Although the force layout pulls connected nodes close to each other, any node with degree $k > 1$ may be connected to any two random nodes in the layout. Since in an ER network

second neighbors are random and uncorrelated, FDL will pull each node with $k > 1$ towards their neighbors, which may be at any random location in the layout. Therefore, we expect the neighbors of a node in ER not to be in the close vicinity of the node, because the neighbors are equally likely to be pulled towards any other node in the layout. As a result, the density of the neighbors of any node i is expected to be uniformly distributed over the whole layout. We will use this observation to calculate the average link length $\overline{R_L}$. The neighbors of each node have uniform spatial density $\rho(x) = L/V_N$. The layout is approximately spherical. Take a node i that is at radius r_i from the center of the layout. The length of a link from node i to a node j at radius r_j is

$$R_{ij} = \sqrt{r_i^2 + r_j^2 - 2r_i r_j \cos\theta_{ij}}, \tag{SI.80}$$

where θ_{ij} is the angle between the vectors from the origin to nodes i and j. To find the average length of links we need to integrate R_{ij} once over its end-node j and then over the position of node i, hence

$$\begin{aligned}\overline{R_L} &= \frac{1}{2}\frac{\int d^3r_i \int d^3r_j R_{ij}}{\int d^3r_i \int d^3r_j} \\ &= \frac{1}{2}\left(\frac{4\pi}{3}R_N^3\right)^{-2}(2\pi)\int^{R_N} d^3r_i \int_0^{R_N} r_j^2 dr_j \int_0^{\pi} \sin\theta d\theta R_{ij}\end{aligned} \tag{SI.81}$$

where the $1/2$ is because r_i and r_j both go over all nodes, thus double-counting the links, the 2π is from the azimuthal integral $d\phi$ in the r_j integral, and the node density drops out because it is constant. The last part of the integral becomes

$$I_0 \equiv \int_{\pi}^{0} d(\cos\theta) R_{ij} = \frac{1}{3r_i r_j}\left(|r_i + r_j|^3 - |r_i - r_j|^3\right) \tag{SI.82}$$

To integrate over r_j, we break the integral down to $r_j \in [0, r_i)$ and $r_j \in [r_i, R_N]$ to get rid of the absolute values, resulting in

$$\begin{aligned}I_1 &\equiv \int_0^{R_N} r_j^2 dr_j I_0 = \int_0^{r_i} r_j^2 dr_j \frac{2}{3r_i}\left(3r_i^2 + r_j^2\right) + \int_{r_i}^{R_N} r_j^2 dr_j \frac{2}{3r_j}\left(r_i^2 + 3r_j^2\right) \\ &= \frac{R_N^4}{2} + \frac{R_N^2 r_i^2}{3} - \frac{r_i^4}{30}.\end{aligned} \tag{SI.83}$$

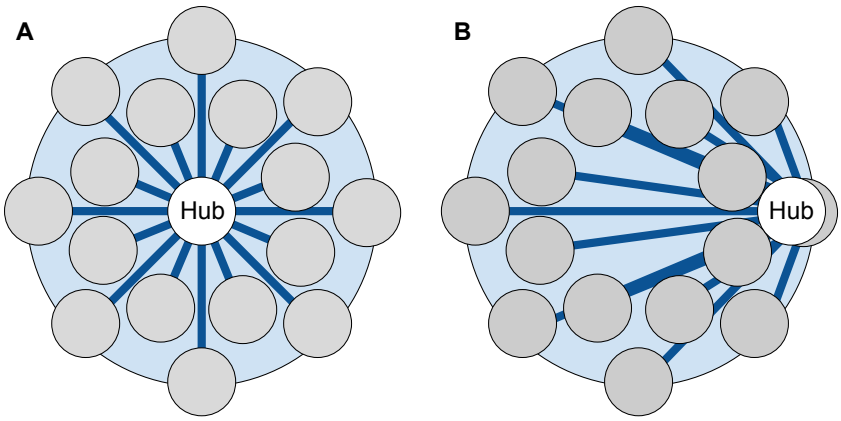

FIG. SI.19. **Location of hubs and value of elastic potential:** In the weak phase, while the repulsive forces derived from V_{NN} and V_{LL} avoid conflicts and are more or less the same for all configurations, the elastic potential V_{el} depends greatly on the arrangement of the nodes. In scale-free networks, hubs have significantly more links than most other nodes. The elastic potential is approximately $V_{el} \approx \frac{k}{2}\sum_i R_{Hi}^2$ in the weak phase, where $R_{Hl} = |\vec{r}_H - \vec{r}_i|$ is the distance between the hib position $\vec{r}_H$ and the position $\vec{r}_i$ of node i. Eq. (SI.86) shows that $V_{el} \propto 3r_H^2 + R_N^2$. Thus, if the hub is placed at the center (A) the total elastic energy will be much smaller than when it is placed off the center (B). Since the optimal layout is effectively minimizing the V_{el}, we can assume that hubs are all close to the center with $r_H \approx 0$.

Finally, the integral over the position $\vec{r}_i$ of node i yields

$$\begin{aligned} I_2 \equiv 4\pi \int_0^{R_N} r_i^2 dr_i I_1 &= \frac{16\pi}{35} R_N^7 \\ \Rightarrow \overline{R_L} = \frac{1}{2}\left(\frac{4\pi}{3}R_N^3\right)^{-2}(2\pi)I_2 &= \frac{18}{35}R_N \end{aligned} \tag{SI.84}$$

Thus, for ER networks the coefficient c_{ER} relating the link length to layout radius $\overline{R_L} = c_{ER}R_N$ is

$$\boxed{c_{ER} = \frac{18}{35} \qquad \text{Erdös-Rényi}} \tag{SI.85}$$

2. Average Link Length in Scale-free Networks

For scale-free networks the considerable heterogeneity in the degrees means the location of nodes within a spherical layout will affect the overall energy cost of the layout in our model. To see this, consider a scale-free network with degree distribution $p(k) \propto k^{-\gamma}$. The most connected nodes, hubs, will have its degree as $k \sim O\left(N^{1/(\gamma-1)}\right)$. For $\gamma = 2$ this means

that this hub will be connected to $O(N)$ of the nodes, or a finite fraction of all nodes. If a hub is placed at radius r_H in an approximately spherical layout, the contribution of the hub to the elastic potential V_{el}^{Hub} will be

$$\begin{aligned} V_{el}^{\text{Hub}} &= \frac{k_{el}}{2}\frac{2\pi k_H}{L}\int_0^{R_N} r_i^2 dr_i \int_\pi^0 d(\cos\theta)\left[r_i^2 + r_H^2 - 2r_i r_H\cos\theta\right] \\ &= \frac{k_{el}}{2}\frac{4\pi k_H}{L}\int_0^{R_N} r_i^2 dr_i \left(r_i^2 + r_H^2\right) \\ &= \frac{k_{el}}{2}\frac{2\pi k_H}{3L} R_N\left(3r_H^2 + R_N^2\right) \end{aligned} \tag{SI.86}$$

where k_{el} is the spring constant and k_H is the hub degree. Eq. (SI.86) is minimized when $r_H = 0$, meaning that the optimal layout will have the biggest hubs near the center of the layout. Since in a scale-free network most of the links are attached to hubs, this means that to estimate the average link length we can approximately assume that all links originate from the center of the sphere, where the hubs are, and remove the r_i integral by putting the density of hubs as $\delta(r_i)$, which results in $R_{ij} = r_j$. This way the average link length becomes

$$\begin{aligned} \overline{R_L} &\approx \frac{\int d^3 r_i \delta^3(\vec{r_i}) \int d^3 r_j R_{ij}}{\int d^3 r_i \delta^3(\vec{r_i}) \int d^3 r_j} \\ &= \left(\frac{4\pi}{3}R_N^3\right)^{-1} \int^{R_N} r_j d^3 r_j = \frac{3}{4}R_N \end{aligned} \tag{SI.87}$$

and the coefficient in $\overline{R_L} = c_{SF} R_N$ becomes

$$\boxed{c_{SF} \approx \frac{3}{4} \qquad \text{Scale-free}} \tag{SI.88}$$

This should be a good approximation for $\gamma \geq 2$ where the hubs are rare, while still accounting for a significant portion of all links. In the extreme case of a star-shaped network (i.e. $\gamma < 2$ where the hub acquires the majority of links) the 3/4 factor becomes exact, as there is only one hub and the optimal layout puts it at the center to minimize the elastic energy. In the other extreme $\gamma \to \infty$ the hubs start disappearing, and all nodes have about the same degree and the situation is similar to ER.

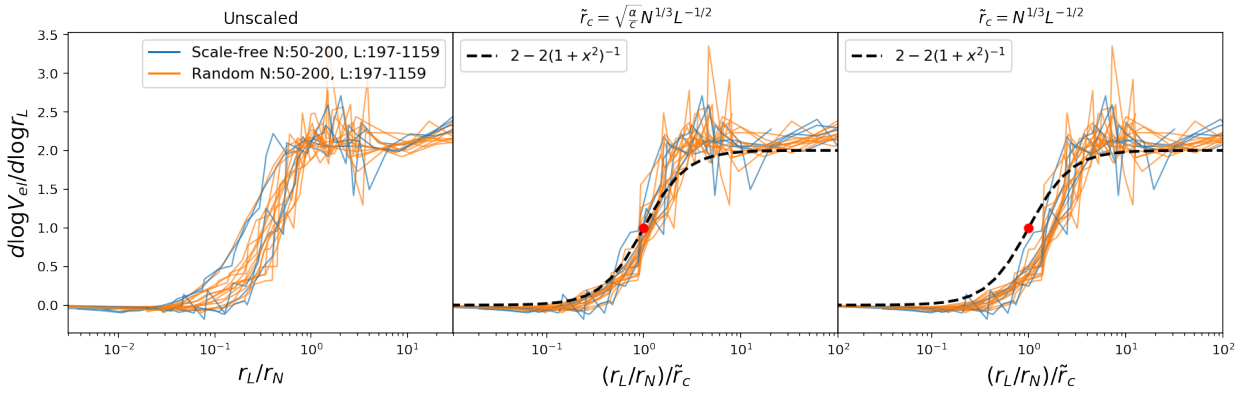

FIG. SI.20. The predicted crossover behavior $r_L^c/r_N \approx N^{1/3}L^{-1/2}$, with and without the prefactor $\sqrt{\alpha/c}$. As we can see, including the prefactor makes the prediction of the transition point much more accurate.

3. Layout Transitions in Lattices and Random Geometric Graphs

The derivation of (SI.58) assumes that links of node i are not only extending to nodes spatially close to i, but also to distant nodes. This assumption holds for random networks, because any node pair has the same chance of being connected, as well as scale-free networks, as most nodes are connected to hubs and thus links connecting to hubs may extend from any part of the layout to the hub. However, it fails for lattices and random geometric graphs (RGG), meaning networks that are formed inside a space via a local rule connecting nodes that are close to each other. These networks are composed of identical sub-networks (such as a unit cell in the lattice), or nearly identical sub-networks (as in RGG). In these cases, N, L and $\overline{R_L}$ in (SI.58) need to be replaced by their corresponding values for the repeated sub-network. Also, increasing the size of these network won't change the location of the transition (Fig. SI.21), because it just increases the number of sub-units, which only interact with each other at their boundaries. This implies that lattices with the same crystalline structure have the same transition point, regardless of their size (for very small lattices, boundary effects may slightly shift this). Fig. SI.21, bottom shows the unscaled curves of the order parameter $\phi(r_L) = d\log\langle l\rangle / d\log r_L$ for regular 3D lattices of various sizes, confirming that they all have the same transition point, as well as the same width for the transition region.

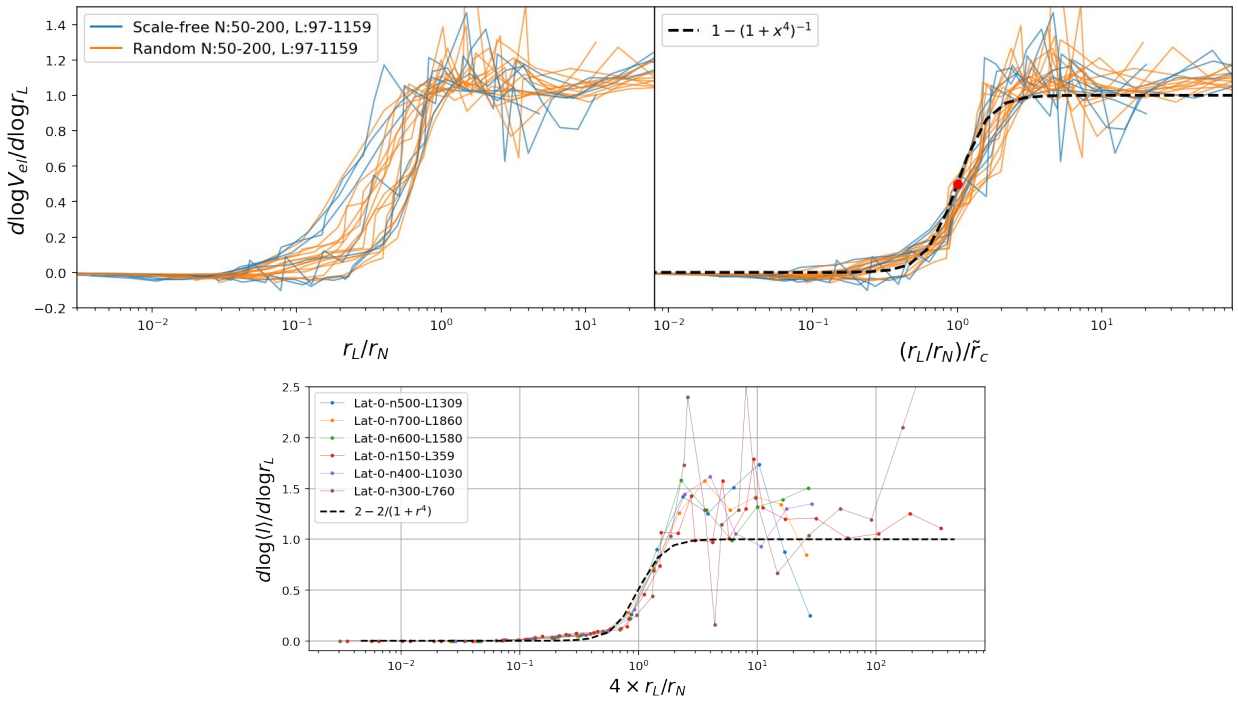

FIG. SI.21. **Transition in Various Topologies:** The quantity $\phi(r_L) \equiv d\log\langle l\rangle / d\log r_L$ distinguishes well between the two regimes ($\langle l\rangle$ being the average link length and keeping r_N constant.) $\phi(r_L)$ is zero deep in the weak regime and it approaches 1 deep in the strongly interacting regime. While scale-free or random networks of various sizes network result in transitions occurring at different values of r_L/r_N (Top, left), scaling by the predicted transition point $\tilde{r}^c = r_L^c/r_N$ collapses the curves well onto a single curve (Top, right). Lattice (Bottom) behave differently than random and scale-free networks. Lattice of all sizes have the same transition point, because once the unit cell has a crossover the whole lattice udergoes a transition.

E. Location and properties of the roots

We want to understand the properties of the roots of (SI.57) to see which of its three roots is relevant to the weak-strong transition. Eq. (SI.57) has the form

$$\rho^3 - a\rho + b = 0 \qquad\qquad a = \frac{c'}{\alpha}N^{-2/3}L, \quad b = \left\langle \frac{k^{3/2}}{3\alpha} \right\rangle \tag{SI.89}$$

with $\rho \equiv r_N/r_L$. Since the coefficient of ρ^2 is zero, the sum of its roots is zero, meaning that it has at least one positive and one negative root. Also, since the product of the three roots is $-b$ and since here $b = \left\langle k^{3/2}\right\rangle/3 > 0$, we definitely have two positive roots and one negative root. Finally, $-b = \rho_-(\rho_+ + \rho_0) + \rho_0\rho_+ < 0$, with (ρ_-, ρ_0, ρ_+) being the three roots in increasing order ($\rho_- < 0$). This results in

$$\frac{1}{\rho_0} + \frac{1}{\rho_+} > \frac{1}{-\rho_-}$$

Since each node must have at least one link, the transition must occur at $r_N \geq r_L$, as multiple links may be fitted around a single node and before r_L becomes as large as r_N, the links will collide. Thus, only the roots $\rho > 1$ are acceptable as a transition point. The root $\rho_+(k)$ can be written out explicitly

$$\rho_+(k) = \frac{A^{\frac{2}{3}} + 12a}{6A^{\frac{1}{3}}}, \qquad A = -12\left(9b + \sqrt{(9b)^2 - 12a^3}\right) \tag{SI.90}$$

with a, b given in (SI.89).

SI 11. NATURE OF THE TRANSITION

Above, we argued that below $\tilde{r}^c$ the layout is dominated by node repulsion (weak regime) and above it by link repulsion (strong regime). Here, we wish to understand the nature of this transition: is it best captured by a first or second order phase transition, or is it best described by a crossover or glass transition? The hallmark of a phase transition is a discontinuity in one of the derivatives of the free energy. Phase transitions are defined in the thermodynamic limit, where the number of degrees of freedom goes to infinity. To establish whether or not in a finite system, such as our networks, the observed change in behavior constitutes a phase transition, a crossover or a glass transition, we must first perform a finite size scaling analysis [9, 37]. Two large classes of behavior are expected:

- *Phase Transition:* If around the transition point some derivatives of the free energy, or other order parameters, diverge when the system size is taken to infinity, that signals the existence of a phase transition. The phase transition point should also become sharp in the infinite size limit.

- *Crossover and Glass Transition:* Lacking a diverging behavior and sharpening of transition point with increased system size, the change in layout should be classified as a crossover, representing a smooth change in behaviour from the weakly interacting to the strongly interacting regime in the thermodynamic limit.

There exist a number of different smooth transitions classified as crossovers [37]. The kind that seems relevant to our model is when a system switches from a mean-field behavior to

a strongly interacting regime where fluctuations in the order parameter become dominant [8]. As we elaborate below, we find that the weak regime of our model can be described by a mean-field theory whereas the strong regime cannot. We have also established that the layouts in the strong regime obey clear scaling laws. Indeed, the order parameter that clearly shows this change from the weak to the strong regime around $\tilde{r}^c$ is the scaling exponent $\phi \equiv d\log\langle l\rangle / d\log r_L$, where $\langle l\rangle$ is the average link length.

A. Finite-size Scaling

We first check how the behaviour of the order parameter $\phi \equiv d\log\langle l\rangle / d\log r_L$ depends on the network size. In particular, we wish to find out whether or not the smooth transition seen in small networks becomes sharper as the network size is increased.

1. Scaling with N and L

To see if the predicted transition point behavior $\tilde{r}^c \propto N^{1/3}L^{-1/2}$ is accurate, we test it on networks with fixed number of links $L \approx 1000$ and varying number of nodes (Fig. SI.22, top) and on networks with a fixed number of nodes, $N = 50$, but varying number of links (Fig. SI.22, bottom). We verified that the curves of $\phi = d\log\langle l\rangle / d\log r_L$ indeed collapse well when r_L/r_N is scaled by $\tilde{r}^c$ (Fig. SI.22).

In addition to verifying the location of the transition, we need to explore how the width of the transition scales with N, L. As we see in Fig. SI.22, after rescaling by the transition point $\tilde{r}^c$, curves of the order parameter $\phi = d\log\langle l\rangle / d\log r_L$ for various N, L collapse onto a single curve. This indicates that, the width *does not* decrease for larger L, N. In particular, this means that the transition region does not become a sharp transition point as N, L are taken to infinity. This is a strong indication that the change of behavior from the weak regime to the strong regime represents a crossover behavior between a non-interacting to a strongly coupled regime.

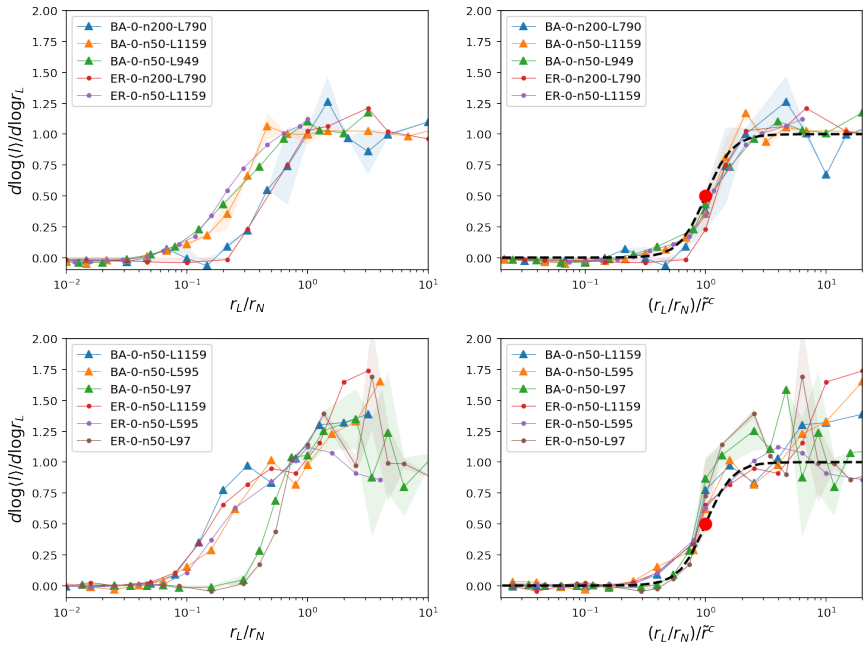

FIG. SI.22. Behavior of the transition point for varying N (top) and L (bottom) separately. The plots on the left and right column are before and after scaling by $\tilde{r}^c$, respectively. The rescaling results in a good collapse of the curves.

2. *Effect of other Parameters of the Potential*

In addition to finite-size effects arising from the network size, we must make sure that the smoothness of the order parameter $\phi = d\log\langle l\rangle / d\log r_L$ in the weak-strong transition region does not arise from the choice of the model parameters.

Effect of the $k/A_{L,N}$ ratio: The model in Eq. (1) is a smooth approximation of a system with hard-wall potentials for the repulsive forces. As such, the transition point exhibits some dependence on the parameters of the potential, such as the ratio of the spring constant to the repulsion amplitude, $k/A_{L,N}$ in links and nodes. This is because, with a smooth potential, the average distance between adjacent nodes or links depends on the magnitude of forces between them. For instance, in the weakly interacting regime, adjacent nodes will not be exactly $2r_N$ apart from each other, and the exact distance depends on the k/A_N ratio. For the node-node repulsion between two adjacent nodes i and j with distance r_{ij} from each other we have

$$|F_{NN}| = A_N \left(\frac{r_{ij}}{2r_N}\right)^{p-1} \exp[\left[-\left(\frac{r_{ij}}{2r_N}\right)^{p}\right] \tag{SI.91}$$

In a regular lattice, this force is offset by the elastic force $2r_{ij}k$ in the links in the weakly interacting regime. Thus, the equilibrium distance between adjacent nodes is found by solving ($y \equiv r_{ij}/(2r_N)$

$$\frac{4r_N k}{A_N} = y^{2-p} e^{-y^p} \tag{SI.92}$$

For example, if we choose $p = 2$, $k/A_N = 0.1$ and $r_N = 1$ the equilibrium distance satisfies

$$\frac{r_{ij}}{2r_N} = \sqrt{-\log \frac{4r_N k}{A_N}} \approx 0.96$$

This would affect the location of the transition point. Unlike the node distance, which is directly affected by the elastic forces in the links, the average link distance is only affected deep in the strongly interacting regime. Thus, the choice of k/A_N should be such that it results in $r_{ij} \approx 2r_N$.

The shift in the transition point under varying k/A_N stems from the average distance between adjacent nodes being a function of k/A_L ratio. Indeed, examining the lattice in the weak regime reveals that the average size of the network changes with the k/A_L ratio. Scaling the r_L/r_N ratio by how much the size of the layout has changed results in an almost perfect collapse of the order parameter for the lattice. To be precise, the "effective" scaling of the network size can be found by measuring the ratio of the average link length to twice the link thickness

$$\frac{r_L^{eff}}{r_L} = \frac{\langle l(r_L) \rangle}{2r_L}$$

where r_L is deep in the strong regime.

Effect of power in the repulsive potential: There doesn't seem to be much dependence on the steepness of the repulsive potential through the exponent p in the $V_{NN} = \sum_{n,m} A_N \exp[|\Delta x_{mn}/2r_N|^p$ (Fig. SI.25).

Number of Link Segments: The numerical simulations are done by discretizing the links into a number of interlaced, interacting beads. This approximation will introduce finite-size effects and the measured transition point does have some dependence on the number of link segments (Fig. SI.26).

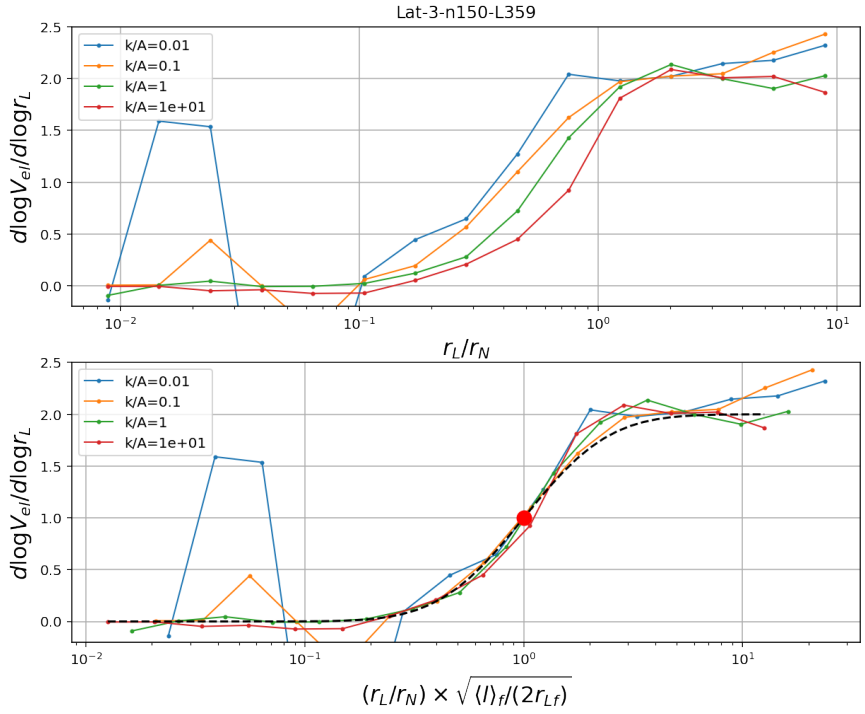

FIG. SI.23. **Collapsing the k/A_L curves for lattice:** the order parameter $d\log\langle V_{el}\rangle/d\log r_L$ shows different transition points for the same lattice with different values of k/A_L. However, changing k/A_L also changes the layout size, which indicates that instead of r_L the layouts have an "effective" $r_L^{eff}(k/A_L)$. This r_L^{eff} scales with $\langle l\rangle^{1/2}$ and rescaling r_L by this growth ratio collapses the transition curves onto a standard error function.

B. The Nature of the Weak-Strong Transition

As we have shown in Fig. SI.22 and 2J,K, after rescaling r_L/r_N by $\tilde{r}^c$, the width of the smooth transition becomes the same for all networks across two orders of magnitude variation in N and L. Additionally, as Figs. SI.23,SI.24,SI.25 and SI.26 indicate, the smoothness of the transition is not due to a lucky choice of various model parameters, such as the elasticity of links, the repulsion amplitudes, steepness of the repulsive potential and number of discrete segments on the links. As a result we conclude that the weak-strong transition represents a crossover from a non-interacting to a strongly interacting phase. There are many types of crossover phenomena [8, 37] in which the system smoothly transitions from one behavior to another. Next we explore their relevance to our system.

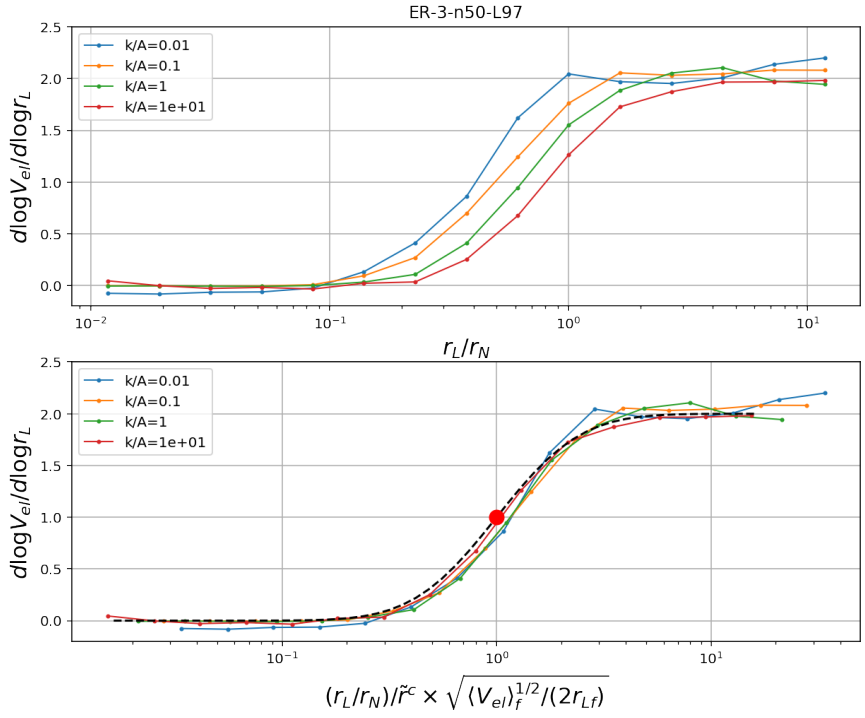

FIG. SI.24. **Collapsing the k/A_L curves for ER:** the order parameter $d\log\langle V_{el}\rangle/d\log r_L$ for different values of k/A_L collapses slightly differently from the lattice. The effective r_L^{eff} scales with the elastic potential $\langle V_{el}\rangle^{1/4}$ and rescaling r_L by this growth ratio and $\tilde{r}^c$ collapses the transition curves onto a standard error function.

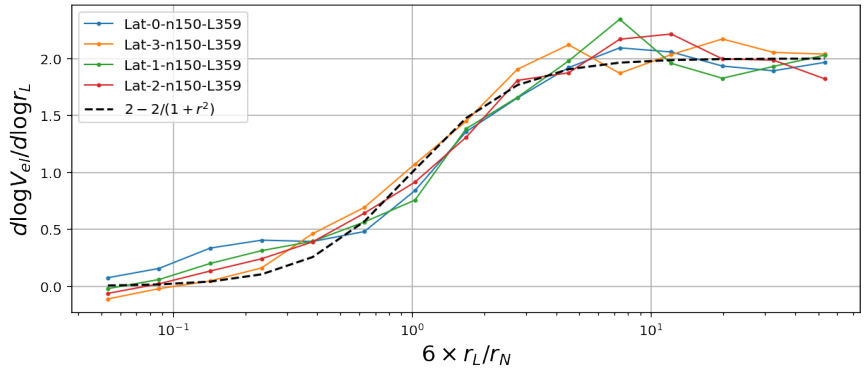

FIG. SI.25. **The effect of node potential steepness:** Simulation of the same lattice with different powers p in the $V_{NN} = \sum_{n,m} A_N \exp\left[|\Delta x_{mn}/2r_N|^p\right]$. The values of p were $2, 3, 4, 6$ for Lat-0,Lat-1,Lat-2,Lat-3, respectively. As we see, there is no significant effect on the transition point arising from p.

1. Weak-strong transition as a crossover

Crossovers are generally discussed in the context of renormalization group (RG) flows [9], where the system has multiple fixed points, with distinct scaling near each one. When the parameters of the system are varied smoothly, they may move the system from the vicinity of one fixed point to another, resulting in a "crosseover" from one scaling behavior to the

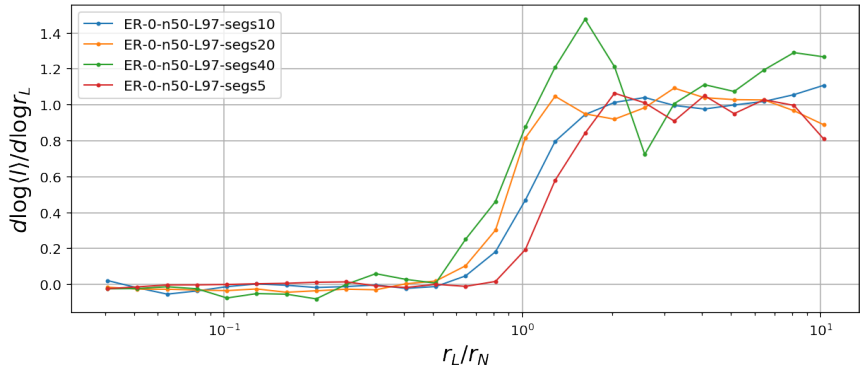

FIG. SI.26. The effect of number of discrete link segments used in simulations. Increasing the number of segments slightly reduces the with of the transition region, and makes the transition slightly sharper.

other.

A subclass of the crossover described above occurs when one of the fixed points corresponds to a non-interacting or weakly interacting regime, well described by a mean-field theory, and the other fixed point corresponds to a strongly interacting regime where the mean-field approximation breaks down [8, 9]. In our model, in the weak regime, the interactions between the links are weak and rare. Thus, the system in the weak regime is well described by a mean-field theory, namely FDL which ignores link crossings. Since links rarely cross each other in the weak regime, the effect of the crossing can be handled with perturbation theory applied to FDL. This is true for both ELI and FUEL. In ELI the nodes are fixed and so the perturbation caused by a link crossing will cause local bending on the links, increasing the overall curvature of the links. In FUEL, the perturbation from link crossing can be compensated for by both bending the link and moving the node and therefore, hence the total link curvature increases by a smaller amount compared to ELI (Fig. 2).

In the strong regime, on the other hand, link interaction is so prevalent that it cannot be handled with perturbation theory. Links interact strongly all along their length with multiple other links in this regime, their trajectories deviating significantly from the non-interacting, straight trajectories of FDL. Therefore, the strongly interacting regime represents a strongly coupled system that can not be described by the mean-field FDL. The strong regime exhibits a clear and distinct scaling behavior (Fig. 2B,C), with the average link length scaling linearly with $\tilde{r} = r_L/r_N$, the curvature scaling as $\tilde{r}^{-1}$ and the elastic potential V_{el} as $\tilde{r}^2$ (Fig. SI.20). Since the macroscopic properties like the average link length, $\langle l \rangle$, and the curvature, $\langle C \rangle$, obey these scaling rules, it seems that the layouts in the strong regime at different $\tilde{r}$ are

self-similar.

We conclude that the weak regime is a fixed point at $\tilde{r} \to 0$ (where link exclusion is irrelevant), described by FDL as the mean-field theory, and the strong regime represents another fixed point at $\tilde{r} \to \infty$ (where node exclusion is irrelevant) where the layout obey clear scaling in terms of $\tilde{r}$.

As the finite-size scaling of the previous section showed, the change from the mean-field behaviour of the weak regime to the scaling behavior of the strong regime is a crossover transition. With this interpretation, the quantity $\tilde{r}^c$ of sec. Sect. SI 10 should be interpreted as the equivalent of the Ginzburg Temperature [8], representing the parameter value at which fluctuations in the order parameter become important, beyond which point the mean-field theory is not valid any more.

---

[1] Dubrovin, B., Fomenko, A. & Novikov, S. *Modern geometry methods and applications. Part 1: The geometry and topology of manifolds (translated by RG Burns)* (Springer-Verlag, Graduate Texts in Mathematics 104, 1984).

[2] Des Cloizeaux, J. Lagrangian theory for a self-avoiding random chain. *Physical Review A* **10**, 1665 (1974).

[3] Mézard, M. & Parisi, G. Replica field theory for random manifolds. *Journal de Physique I* **1**, 809–836 (1991).

[4] Gay, J. & Berne, B. Modification of the overlap potential to mimic a linear site–site potential. *The Journal of Chemical Physics* **74**, 3316–3319 (1981).

[5] Bouchaud, J.-P., Cugliandolo, L. F., Kurchan, J. & Mezard, M. Out of equilibrium dynamics in spin-glasses and other glassy systems. *Spin glasses and random fields* 161–223 (1998).

[6] Kirkpatrick, S., Gelatt Jr, C. D. & Vecchi, M. P. Optimization by simulated annealing. In *Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications*, 339–348 (World Scientific, 1987).

[7] Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).

[8] Chaikin, P. M. & Lubensky, T. C. *Principles of condensed matter physics* (Cambridge University Press, 2000).

[9] Cardy, J. *Scaling and renormalization in statistical physics*, vol. 5 (Cambridge university press, 1996).

[10] Irgens, F. *Continuum mechanics* (Springer Science & Business Media, 2008).

[11] Ahn, Y.-Y., Ahnert, S. E., Bagrow, J. P. & Barabási, A.-L. Flavor network and the principles of food pairing. *Scientific Reports* **1** (2011).

[12] Zeeman, E. C. Unknotting combinatorial balls. *Annals of Mathematics* 501–526 (1963).

[13] Stepanyants, A., Hof, P. R. & Chklovskii, D. B. Geometry and structural plasticity of synaptic connectivity. *Neuron* **34**, 275–288 (2002).

[14] Rivera-Alba, M. *et al.* Wiring economy and volume exclusion determine neuronal placement in the drosophila brain. *Current Biology* **21**, 2000–2005 (2011).

[15] Ventura-Antunes, L., Mota, B. & Herculano-Houzel, S. Different scaling of white matter volume, cortical connectivity, and gyrification across rodent and primate brains. *Frontiers in neuroanatomy* **7**, 3 (2013).

[16] Bullmore, E. & Sporns, O. The economy of brain network organization. *Nature Reviews Neuroscience* **13**, 336–349 (2012).

[17] Sporns, O., Chialvo, D. R., Kaiser, M. & Hilgetag, C. C. Organization, development and function of complex brain networks. *Trends in Cognitive Sciences* **8**, 418–425 (2004).

[18] Kasthuri, N., Lichtman, J. *et al.* Saturated reconstruction of a volume of neocortex. *Cell* **162**, 648–661 (2015).

[19] De Gennes, P.-G. *Scaling concepts in polymer physics* (Cornell University Press, 1979).

[20] Davidson, R. & Harel, D. Drawing graphs nicely using simulated annealing. *ACM Transactions on Graphics (TOG)* **15**, 301–331 (1996).

[21] Kamada, T. & Kawai, S. An algorithm for drawing general undirected graphs. *Information Processing Letters* **31**, 7–15 (1989).

[22] Noack, A. Modularity clustering is force-directed layout. *Physical Review E* **79**, 026102 (2009).

[23] Parisi, G. The physical meaning of replica symmetry breaking. *arXiv preprint cond-mat/0205387* (2002).

[24] Pelissetto, A. & Vicari, E. Critical phenomena and renormalization-group theory. *Physics Reports* **368**, 549–727 (2002).

[25] Berne, B. J. & Pechukas, P. Gaussian model potentials for molecular interactions. *The Journal of Chemical Physics* **56**, 4213–4216 (1972).

[26] Everaers, R. & Ejtehadi, M. Interaction potentials for soft and hard ellipsoids. *Physical Review E* **67**, 041710 (2003).

[27] Babadi, M., Everaers, R. & Ejtehadi, M. Coarse-grained interaction potentials for anisotropic molecules. *The Journal of Chemical Physics* **124**, 174708 (2006).

[28] Mergell, B., Ejtehadi, M. R. & Everaers, R. Modeling dna structure, elasticity, and deformations at the base-pair level. *Physical Review E* **68**, 021911 (2003).

[29] Cleaver, D. J., Care, C. M., Allen, M. P. & Neal, M. P. Extension and generalization of the gay-berne potential. *Physical Review E* **54**, 559 (1996).

[30] Landau, L. D. & Lifshitz, E. Theory of elasticity, vol. 7. *Course of Theoretical Physics* **3**, 109 (1986).

[31] Stillinger, F. H. & Weber, T. A. Packing structures and transitions in liquids and solids. *Science(Washington, DC)* **225**, 983–9 (1984).

[32] O'Hern, C. S., Langer, S. A., Liu, A. J. & Nagel, S. R. Random packings of frictionless particles. *Physical Review Letters* **88**, 075507 (2002).

[33] O'Hern, C. S., Silbert, L. E., Liu, A. J. & Nagel, S. R. Jamming at zero temperature and zero applied stress: The epitome of disorder. *Physical Review E* **68**, 011306 (2003).

[34] Abadi, M. *et al.* TensorFlow: Large-scale machine learning on heterogeneous systems (2015). URL `https://www.tensorflow.org/`. Software available from tensorflow.org.

[35] Barabási, A.-L. *Network Science* (Cambridge University Press, 2016).

[36] Pinelis, I. Positive-part moments via characteristic functions, and more general expressions. *Journal of Theoretical Probability* **31**, 527–555 (2018).

[37] Goldenfeld, N. *Lectures on phase transitions and the renormalization group* (CRC Press, 2018).