

SCIENTIFIC REPORTS



OPEN

Multi-scale structure and topological anomaly detection via a new network statistic: The onion decomposition

Laurent Hébert-Dufresne¹, Joshua A. Grochow¹ & Antoine Allard²

Received: 03 May 2016

Accepted: 25 July 2016

Published: 18 August 2016

We introduce a network statistic that measures structural properties at the micro-, meso-, and macroscopic scales, while still being easy to compute and interpretable at a glance. Our statistic, *the onion spectrum*, is based on the *onion decomposition*, which refines the *k*-core decomposition, a standard network fingerprinting method. The onion spectrum is exactly as easy to compute as the *k*-cores: It is based on the stages at which each vertex gets removed from a graph in the standard algorithm for computing the *k*-cores. Yet, the onion spectrum reveals much more information about a network, and at multiple scales; for example, it can be used to quantify node heterogeneity, degree correlations, centrality, and tree- or lattice-likeness. Furthermore, unlike the *k*-core decomposition, the combined degree-onion spectrum immediately gives a clear local picture of the network around each node which allows the detection of interesting subgraphs whose topological structure differs from the global network organization. This local description can also be leveraged to easily generate samples from the ensemble of networks with a given joint degree-onion distribution. We demonstrate the utility of the onion spectrum for understanding both static and dynamic properties on several standard graph models and on many real-world networks.

The ever-growing literature on complex networks is a testament both to the ubiquity of networks as conceptual tools, and to the lack of a definitive toolbox. The recent explosion in available data, both in the variety of sources (e.g., social networks, biological networks, infrastructure or technological systems) and in the sheer size of the datasets, stresses the need for good methods to analyze and synthesize information about the structure of networks.

Although there is already a plethora of metrics and methods to study networks¹, there is still a need for multi-scale metrics. For example, degree distribution² and local clustering³ are incredibly simple and informative but only capture microscopic, local features of networks. Modularity^{4,5} and other community structure properties probe the meso-scale organization of networks but are still ill-defined, as evidenced by the lack of a common definition for network communities⁶. Macroscopic measures such as betweenness centrality⁷, eigenvector centrality¹, and mean shortest path length³ can characterize the role of a given node in the overall network structure. However they demand significant computational effort and rarely map back to a good local understanding of how the network might be constructed. There is a dire need for tools that are easy to compute and that complement existing tools by characterizing networks at multiple scales at a glance.

As it turns out, the answer was perhaps hidden all along in a popular algorithm: The *k*-core decomposition^{8,9}. We refine this algorithm both to characterize networks as a whole and to detect some of their interesting, unique subgraphs. We will here focus on undirected and unweighted networks, but our contribution could potentially be extended to existing generalizations of the *k*-core decomposition^{10–15}. We will discuss how our method can be described in terms of local rules, yielding the most structurally constraining network connection process to date. Finally, we show how these constrained networks can be used to better capture a dynamical process than existing random network models.

¹Santa Fe Institute, Santa Fe, NM 87501, USA. ²Departament de Física de la Matèria Condensada, Universitat de Barcelona, Barcelona 08028, Spain. Correspondence and requests for materials should be addressed to L.H.-D. (email: laurent@santafe.edu)

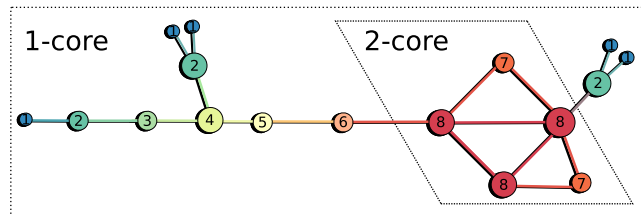


Figure 1. Onion decomposition on a simple graph. Visualization of the OD where node size corresponds to their degree, node color and tag to the layer in which they are removed, and where we have highlighted the corresponding k -cores. The OD first removes all nodes of degree 1 as there are no degree 0 (layer 1) then nodes that are now of degree 1 (layer 2) and so on (layers 3–6) until we reach the 2-core where all nodes have degree at least 2. The OD then starts over, now removing nodes of degree 2 or less (layers 7–8).

The Onion Decomposition

The k -core decomposition is a network pruning method whose goal is to separate a network into a succession of nested cores, effectively defining a center and periphery to the network. The method is based on the concept of k -cores: The maximal sub-network within which every node has degree *at least* k . Nodes that are in the k -core but not in the $(k + 1)$ -core can be said to belong to the k -shell and are thus of *coreness* k . Lower k -shells can be thought of as being less central; the network is then viewed as a series of progressively denser and more central cores.

A useful metaphor relates the k -core decomposition to the peeling of an onion¹⁶: One first removes all nodes of degree 1, then nodes that are now of degree one following the removal of the first onion skin, and so on until all nodes left are of degree greater than one and thus compose the 2-core. The peeling then starts again, now removing nodes of degree at most 2. While the mathematical definition of the k -core is elegant, we argue that much more information lies in the decomposition process than in the final results. In other words, we aim to study how the speed at which one can peel the network into cores is related to its structure.

We thus introduce the concept of *layers*: How many peeling passes are needed to reach a given node. For instance, nodes of the k -shell belong to its first local layer if they are of degree exactly k within the k -core, or to its second local layer if they are of degree at most k only after the removal of the first layer.

The procedure to identify these layers, which we call the *onion decomposition* (OD), is essentially the same as that for the k -core decomposition, but retains more information. We generalize the algorithm for k -cores⁹ with a very simple modification to produce the OD (see Fig. 1 for a schematic):

1: Input graph with vertex set V and neighbors N
2: Output Onion <i>Layer</i> and <i>Coreness</i> of each vertex
3: D : = list of degrees
4: $core$: = 1; $layer$: = 1
5: while V is not empty do
6: ThisLayer: = $\{v \in V \mid D(v) \leq core\}$
7: for each $v \in$ ThisLayer do
8: $Coreness(v)$: = $core$; $Layer(v)$: = $layer$
9: for each $w \in N(v)$, decrease $D(w)$ by 1 end for
10: Remove v from V and D
11: end for
12: $layer$: = $layer + 1$
13: if the minimum degree in D is $\geq (core + 1)$ then
14: $core$: = $\min D$;
15: end if
16: end while

The run-time of this algorithm scales as $O(|E| \log |V|)$. For the run-time analysis of the OD, as well as implementation notes that can affect the speed of the algorithm in practice, see Supplementary Information.

The OD thus provides a centrality measure more refined than coreness, and essentially at least as easy to compute as other centrality measures. For example, the best known algorithm for betweenness centrality takes time $O(|V||E| + |V|^2 \log |V|)$ ¹⁷, and even computing eigenvector centrality takes time $O(|E|/\delta)$ (where δ is the gap between the first and second eigenvalues), which is comparable to the OD. We also show how the OD naturally corresponds to an ensemble of random networks, unlike other centrality measures.

Results

With this new analysis method in hand, we define the *onion spectrum* of a network as the fraction of all nodes which are found in a given layer of the OD. The onion spectrum can be thought of as a structural spectrum as it assigns every node to a given structural role through our new measure of node centrality. As we will see, even a

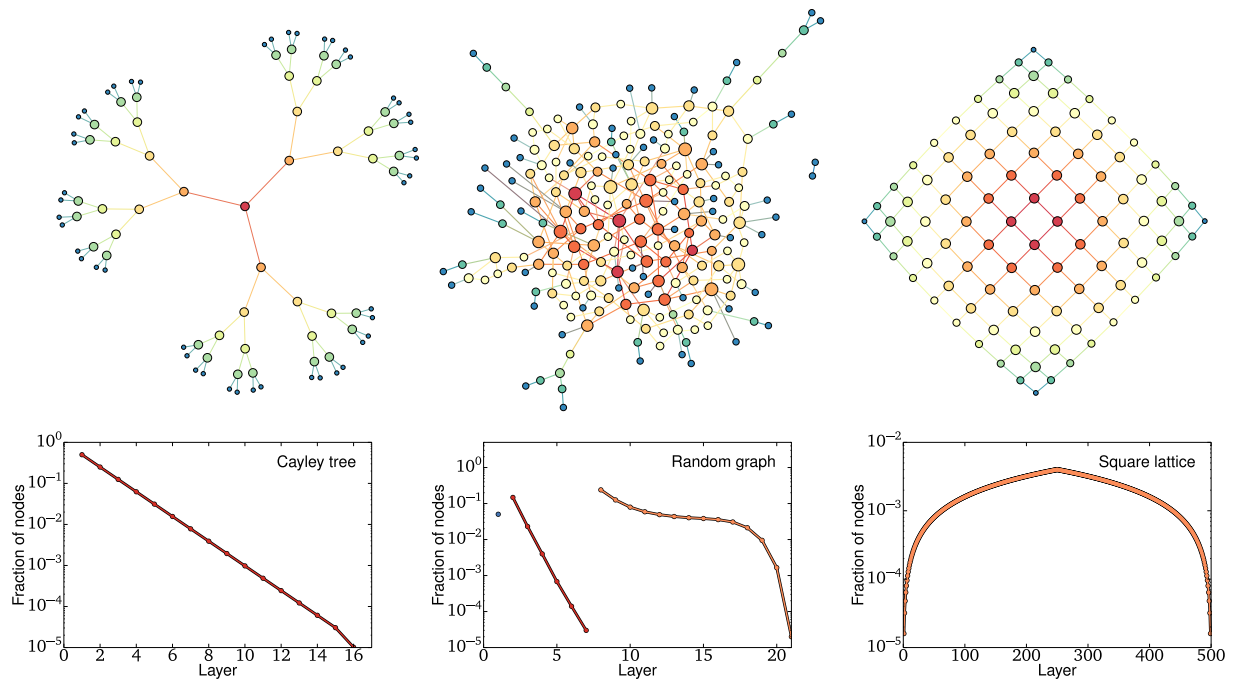


Figure 2. A few model networks and their onion spectra. (left) The perfect tree with branching factor of two. (middle) A fully random Erdős-Rényi graph. (right) A square lattice. In all cases, node size is proportional to its degree within the shown subset while node color indicates where the nodes are found in the onion decomposition, from blue for shallow nodes to red for deep nodes. In the onion spectra, dot colors and lines are used to indicate layers corresponding to the same k -shell. Note that colors are used for contrast and do not correspond to node color in the network cartoons. Both the Cayley tree and the square lattice only have one shell, corresponding to a 1-core and a 2-core, respectively, while the random graph features 3 cores, roughly corresponding to disconnected nodes, a tree-like periphery and a dense core. The distribution of nodes per layer in the Onion Decomposition tell us something about the structure as, for example, a tree is explored exponentially and a lattice linearly.

glance at the onion spectrum of different networks provides significant insights into their structure at multiple scales.

Model networks. We first test the OD on model networks designed to gather insights on the types of onion spectra produced by different network structures: a perfect Cayley tree with a branching factor of 2, an Erdős-Rényi graph with fixed density and a square lattice. Small versions of these networks are presented in Fig. 2. Interestingly, the k -core decomposition provides very little information on both the Cayley tree and the square lattice. Indeed, all nodes belong to the same core in both graphs despite the fact they feature a well-defined center and periphery. These networks are used to test the behavior of the OD on different structures: tree-like branching, core-periphery structure and “geographic-like” embedding. We show the results of the OD on these model networks in Fig. 2. Certain features of their onion spectra can also be written analytically (see Supplementary Information).

In the case of the Cayley tree, the nodes of the original network have a very well-defined position in the structure given by their distance to the leaf nodes. After removal of the first onion layer, the network is essentially the same (with one fewer layer) and the distances between all nodes and the leaves are reduced by one. The process then goes on, such that all nodes are within the 1-core, and we can find an exponentially decreasing number of nodes in all layers (i.e., as an inverse branching process). In fact, all trees feature an exponential onion spectrum where the ratio of two subsequent layer densities roughly corresponds to an average branching factor. Locally speaking, this tree-like property is observed in most sparse networks.

In the Erdős-Rényi graph, the first few layers are peeled in a similar fashion to a perfect tree. This can be observed in Fig. 2 (top center) where nodes of degree one (in blue) are connected to peripheral nodes of higher degree (in green). These various branches define the periphery of the network and comprise the first cores of the decomposition. The last core, much denser, features a significant number of loops on all scales. The density of onion layers within that core thus follow a very different decay, clearly sub-exponential until the finite size cut-off. To better understand this behavior, we turn to the square lattice, which possesses loops of all possible even lengths, by design.

To analyze the square lattice, it is useful to go back to Fig. 2 (top right) and notice how the network is explored by the OD. The first nodes to be removed are the corners, then their neighbor, and so on. The sites of nodes removed in each layer follow horizontal/vertical lines converging to the center with increasing layers. The unique

Scale	Signature in the spectrum	Property
micro	more cores than random	assortativity
	fewer cores than random	disassortativity
meso	exponential decay	tree-like structure
	sub-exponential decay	loopy structure
	change in decay	interesting subgraph
macro	cores	core-periphery structure
	layers	node centrality

Table 1. Summary of some properties found in the onion spectrum. “Randomness” here refers to rewiring preserving only degree distribution. All other properties can also be mapped to local connection rules using our network reconstruction rules.

shape of the onion spectrum is easy to calculate and is found to be linear, both in the increasing and decreasing regime (see Supplementary Information).

To broadly summarize the conclusions that can be drawn from the three model networks: The onion spectrum of tree-like and sparse random networks are expected to decay exponentially, whereas those of lattices and dense subgraphs fall sub-exponentially (see Table 1).

Real-world case studies. We now apply the OD to several real-world networks. These networks are selected to cover a few structural features that are captured by the OD so as to highlight its use as a macroscopic, mesoscopic and microscopic tool (many more are explored in the Supplementary Information). Namely, we investigate how the onion spectrum allows us to distinguish tree-like and lattice-like networks, how it allows us to identify topologically anomalous subgraphs whose local structure differs from the global organization, and how the degree distribution and degree-degree correlations are reflected in a network’s onion spectrum (see Table 1).

Figure 3 presents the onion spectra of two infrastructure networks: The Northwestern American power grid³ and the Pennsylvania road system¹⁸. In both cases, the figure presents a subset of the network that reflects its global structure. The selected layers are shaded in the spectrum, and parts of those subgraphs are presented in the top half of the figure. Once again, merely glancing at the decay of layer density informs us as to how these networks occupy space: The exponential decay of the power grid’s onion spectrum is a signature of its (effectively) tree-like structure, where most loops are short and link nearby nodes. This is in stark contrast with the sub-exponential spectrum and structure of the road network, where despite featuring fewer triangles, there are loops occurring on all scales, as observed in Fig. 3(top right). To visualize the difference, note that removing a link or two can disconnect many nodes from the power grid, but not on the road network.

While the major cores of both the power grid and the road network have a very steady behavior, one can look for significant deviation from general trends—i.e., topological anomalies—to identify interesting subgraphs that stand out from the overall network structure. To this end, we present the onion spectrum of a co-authorship network¹⁹ in Fig. 4(left). Like most sparse social networks, the overall structure is roughly tree-like. However, the onion spectrum allows us to identify interesting subgraphs, leading us to focus on the nodes around layer 175 (highlighted), as we see a radically different decay there. This anomalous subgraph is composed of nodes with degree at least 18 that, once peripheral nodes are removed, are found to be organized in communities of long-time, prolific collaborators. Similar topological anomalies, with similar explanations, are seen in other co-authorship networks (see Supplementary Information). This locally high modularity contrasts with the global tree-like structure, and is easily picked up by the OD. A subset of this subgraph is presented on Fig. 4(right).

The condensed matter co-authorship network also illustrates how the onion spectrum can detect degree correlations. Not only does this network have more cores than the previous networks because of its fat-tailed degree distribution, it even has significantly more cores than would be expected from a network with the same degree distribution but otherwise random connections: 29 versus 9, respectively (see Supplementary Information). This is a signature of positive degree correlations, often called *homophily*²⁰ or assortativity, meaning neighboring nodes tend to have similar degrees²¹. It is intuitively clear that having high-degree nodes share connections with other high-degree nodes favors the emergence of higher coreness. By comparison to a randomized network, one can thus look for degree correlations by comparing how many cores are found within the network (see Supplementary Information for more examples). That being said, the decay rates within cores can also contain further information about the network structure and this is especially true when faced with negative degree correlations.

The onion spectrum of a subset of the World Wide Web, the stanford.edu domain¹⁸, is presented in Fig. 5(middle). In this network, even within a single layer, the decay rate varies significantly. In the Supplementary Information, we show that this variation is a consequence of negative degree-degree correlations, by comparing the network with two ensembles: one in which only the degrees are preserved, and one in which furthermore the degree-degree correlations are preserved. In the former, the decay rates within each layer are roughly constant, whereas in the latter the decay rates closely match the true, observed rates. Essentially, this variation is caused by the fact that most links stemming from low degree nodes are connected to nodes of higher degree than would be expected by chance. This negative correlation, or disassortativity, implies that fewer nodes are removed within the second and subsequent layer of a core than expected. This also implies that the network contains fewer cores than a randomized version. We propose a better rewiring scheme in the next section.

Finally, we use the stanford.edu domain to illustrate how other types of interesting subgraphs can be identified. In the case of most sparse networks, being “interesting” means at least deviating from the expected

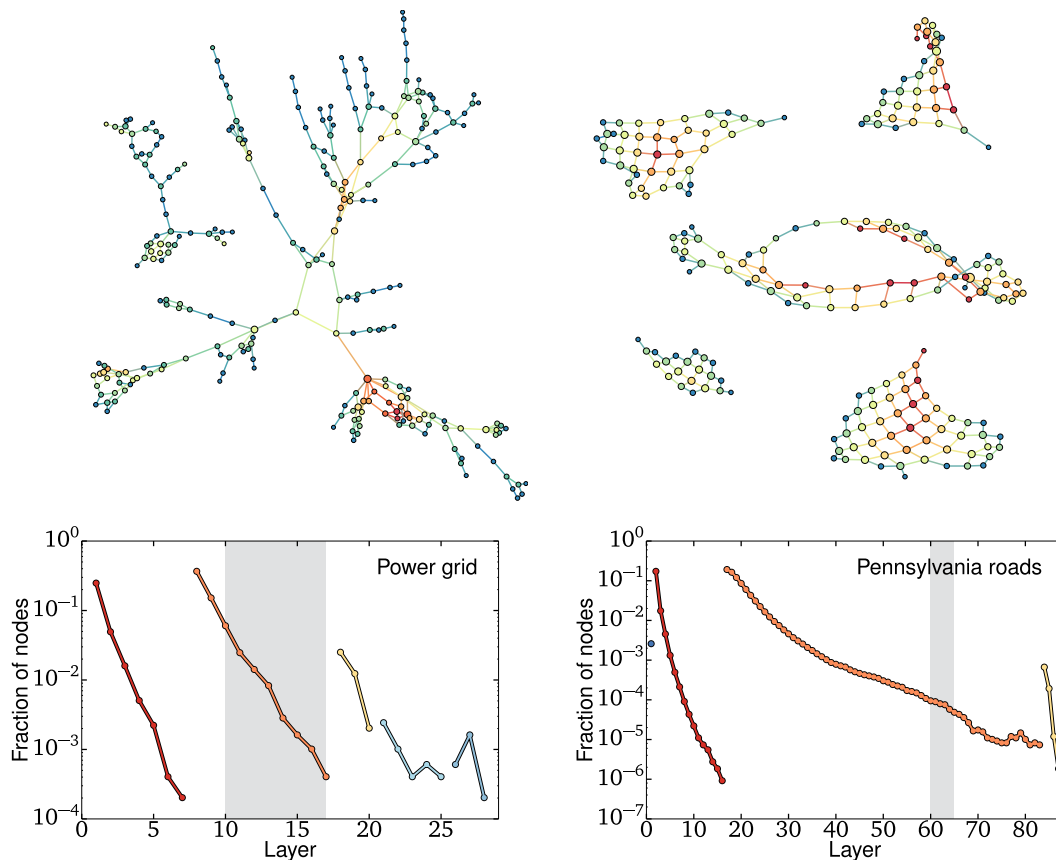


Figure 3. Onion spectra of the Northwestern American power grid and of the Pennsylvania road system and representative subgraphs. As in the model networks of Fig. 2, the tree-like structure of the power grid and the lattice structure of the road networks are reflected in the exponential and sub-exponential decay of layer density per shell in the onion spectra, respectively. Again, dot colors in the spectra reflect layers belonging to different shells whereas node color in the networks correspond to the shells in which they are found (shaded layers in the spectra, only links between nodes belonging to the subgraph are shown).

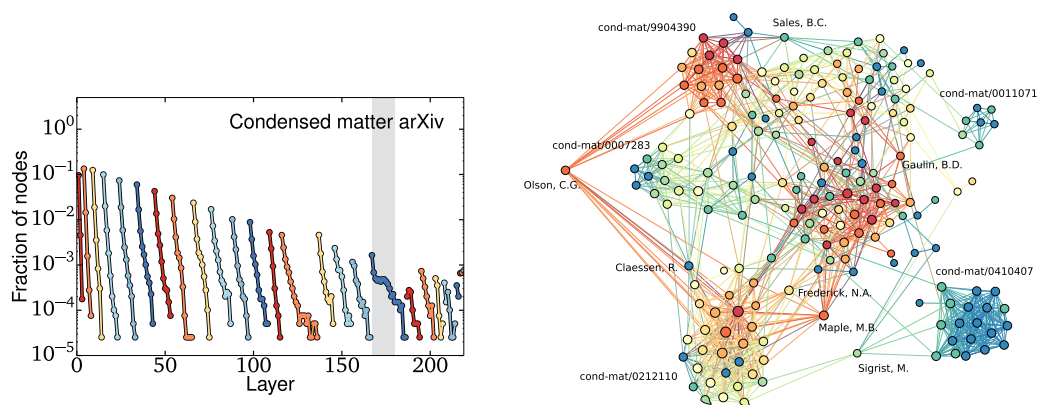


Figure 4. The condensed matter co-authorship onion spectrum and a selected subgraph. While the overall structure of this sparse network is tree-like, some unique subsets can be identified in the spectrum, one of which is shown (shaded layers). Dense communities of collaborators are identified with a selected arXiv identifier and some of the authors bridging communities are identified by name.

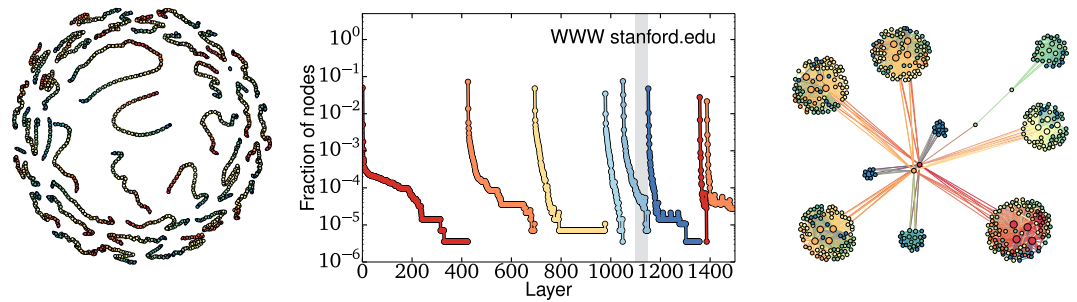


Figure 5. The WWW subset of stanford.edu and its onion spectrum. This final case study highlights how the method can uncover backbones and governing nodes that can only be highlighted once peripheral nodes are removed. In particular, we illustrate this with extremely unlikely chains of nodes in the first core (left) and a centralized community structure in the sixth core (right).

tree-like structure. In Fig. 5, we identify two (out of many) interesting subgraphs that present a significantly sub-exponential decay of density per layer. The first is interesting as it appears in the first core, where nodes are expected to be of low degree such that large non-tree-like subgraphs are unlikely. Yet, through the OD, we have identified a set of around 8500 nodes of degree two joined in chain-like fashion (approximately 3% of the entire network). Based on the degree distribution alone, these subgraphs had a probability less than 10^{-15000} of occurring (i.e., it should not occur). But somehow, such surprising structures seem to appear on all scales within the domain, as similar slow decay can be observed in all cores. For instance, the subgraph shown in Fig. 5(right) occurs within the 5-core and appears to be a centralized community structure where many different groups are governed by only a few nodes. The entire subgraph collapses under the OD once the smallest groups are removed. In the context of World Wide Web networks, we thus believe that this procedure can help uncover the backbone and governing nodes of different domains.

Although other methods might be designed to detect these particular structures—backbones, governing nodes, and the structure in Fig. 4—we point out that the OD allowed us to detect these structures in these networks *de novo*, without knowing what kind of structure we were looking for ahead of time.

The onion network ensemble (ONE). We have shown how the onion decomposition allows to characterize a complex network, at a glance, through the general trends of its onion spectrum. The OD achieves this in part because it aggregates multiple properties of complex networks, at different scales, into a single spectrum: degree heterogeneity and correlations, organization in terms of centrality, and tree-likeness or loop prevalence.

To the best of our knowledge, the onion decomposition is the first method to combine multiple scale of analysis into one algorithm that is simultaneously fast (it scales as $O(|E| \log |V|)$), simple, well-defined, mathematically principled, and allows *de novo* detection of interesting subgraphs as well as network reconstruction. Indeed, and contrary to other methods of network analysis (e.g., community detection algorithms), the positions of the nodes in the OD are elegant and straightforward: (i) a node is in the k -core if it is in the maximal subset of the network in which every node has degree at least k within the subset, and (ii) within the k -core, a node is in the ℓ -th local layer if one must remove all nodes of degree at most k a total of $\ell - 1$ times before reaching the node.

As a final point, we show how the OD naturally lends itself to the definition of the most structurally constraining, yet still easily and exactly sampleable, network randomization procedure to date. We then show how these constrained networks outperform existing random network models at reproducing the outcome of the Susceptible-Infectious-Susceptible dynamical process, as well as at reproducing the full distribution of all-pairs shortest path lengths.

Construction of the ONE. The specifications of the OD can be translated into a simple set of connection rules defining an ensemble of networks that preserve the joint degree-onion spectrum (i.e., the degree and the position in the onion spectrum of nodes). We refer to this ensemble as the onion network ensemble (ONE). Note that using only the k -core structure one only gets a rewiring scheme that *approximately* preserves that structure²², but adding layer information allows us to *exactly* preserve the coreness and layer of every single node. Given a joint degree-onion spectrum, the ONE can be sampled quasi-uniformly, by a model only marginally more complicated than the configuration model. This is in strong contrast with other centrality measures such as betweenness or eigenvector centrality, whose translations into local connection rules, if any, remain open problems.

In the ONE, a node of degree d and coreness k belonging to layer ℓ must be connected to other nodes in the following fashion. If the node is located in the first layer of its core, it must have exactly k links to nodes of layers $\ell' \geq \ell$ since all previous layers are not in the k -core but in the $(k - 1)$ -core. Otherwise, the node must have *at least* $k + 1$ links to nodes of layers $\ell' \geq \ell - 1$ and at most k links to nodes of layers $\ell' \geq \ell$. Links to layer $\ell - 1$ anchor the node in the correct layer and those to layers beyond provide the node with the correct coreness. Rewiring links according to these rules while enforcing the correlations between layers²² allows to explore the ONE associated to a given original network, while simultaneously preserving its degree distribution and its onion spectrum. This rewiring process provides a structurally constrained null model against which real networks can be compared, and that can also be used to build networks of *different sizes* with the same joint degree-onion spectrum. This allows comparison of graphs across sizes, as well as analysis in the limit of very large networks. Note that in this limit, the ONE exactly defines a network without loops or correlations (other than the ones between layers that are explicitly enforced).

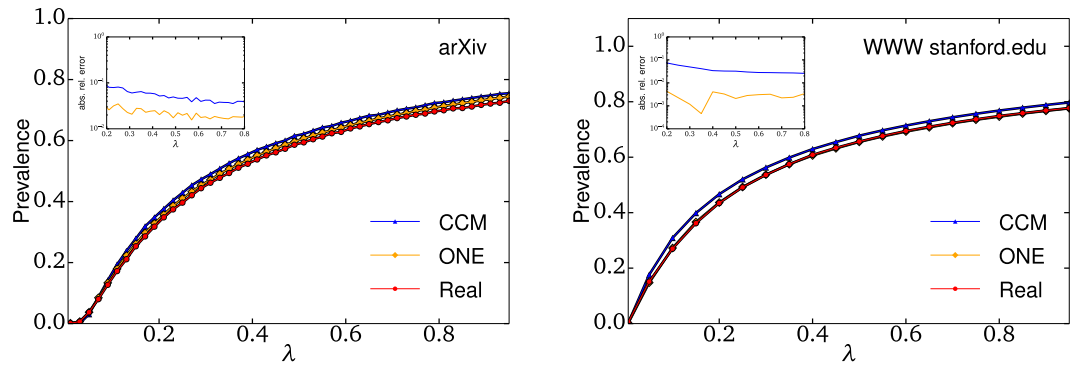


Figure 6. Simulations of the SIS epidemic process. The prevalence (number of infectious nodes) associated with the steady state of the SIS dynamics as a function of the dimensionless parameter λ on (left) the co-authorship network from the arXiv and (right) links between stanford.edu pages. Simulations on the real original networks are shown in red, and those on networks rewired according to the Correlated Configuration Model (CCM) and the ONE are shown in blue and orange, respectively. The insets show the absolute value of the relative errors on predictions from the rewired networks. Note that prevalence observed on the stanford.edu network and on its ONE are almost indistinguishable. In addition to its more accurate results, the ONE requires significantly less information than the CCM.

Dynamical process on the ONE. Let us consider the Susceptible-Infectious-Susceptible model of disease spread as an illustration of the general utility of the OD—and more specifically of the ONE—as an effective network structure. In this dynamics, nodes are either susceptible or infectious, infectious nodes infect their susceptible neighbors at a rate β and they recover at a rate α . The SIS dynamics is governed by the dimensionless parameter $\lambda = \beta/\alpha$ and, for λ above a critical value, possesses a steady state in which a non-zero fraction of the nodes are infectious²³.

We simulate this process on two real networks on which diseases, ideas or computer viruses could be expected to spread. Figure 6 gives the prevalence of the SIS dynamics as a function of λ and compares it with the one observed on networks rewired according to the correlated configuration model (CCM) and the ONE. The CCM produces networks that respect a given degree distribution as well as the entire degree-degree correlation matrix (i.e., probability that a randomly selected link connects nodes of degree d_i and d_j). The most striking result of Fig. 6 is that the ONE outperforms the CCM even though it requires *less* information. Indeed, the CCM requires a joint degree-degree correlation matrix which scales as k_{\max}^2 , whereas the ONE requires information that scales as $k_{\max} \ell_{\max} \sim k_{\max}^{3/2}$ (see Supplementary Information). This comparison with the CCM reveals that the performance of the ONE is not merely a consequence of the degree-degree correlations encoded in the layer-layer correlations, but stems from the global features encoded in the OD and in the ONE's local connection rules.

It is worth noting that the SIS dynamics is highly self-averaging, as the infection constantly moves and revisits every node. Consequently, the SIS model does not depend much on the high-level structure of a network, but more on local features. Considering how much *local* information is needed by the CCM, it is somewhat surprising that the ONE also performs better than the CCM in an application that depends mostly on local information.

Structure of the ONE. An interesting feature of our SIS results presented in Fig. 6 is that the ONE manages to provide precise approximations despite being a tree-like random network model. Indeed, it is often thought that models like the CCM overestimate the outcome of a spreading process mostly because they lack the clustering (or modularity) of real-world networks. However we saw how loops, which are the signature of clustering and modularity, affect the onion spectrum of a network. In extreme cases, such as going from a perfect tree to a lattice, loops can change the exponential decay of the spectrum to a linear function. Thus, despite being a tree-like model, the ONE *stretches* the network by considering layers that somehow encode those loops. In other words, while the connection rules of the ONE are strictly local, they mimic the global organization of a given network. For instance, and in contrast to standard rewiring algorithms, the ONE preserves well the diameter of a network, and even offers a good approximation of the entire distribution of shortest path lengths (see Fig. 7 and Supplementary Information). Again, the ONE outperforms the CCM while requiring less information.

Discussion

In summary, we have introduced the onion decomposition that can be useful to (i) characterize complex networks, (ii) identify interesting subgraphs and (iii) approximate their structure—and the dynamics they support—through the onion network ensemble.

This is done by running an algorithm that scales almost linearly with network size ($O(|E| \log |V|)$) to reveal how quickly a network can be peeled by removing peripheral nodes. The density, or number of nodes, in every layer of the peeling process is called the onion spectrum and characterizes both the core-periphery structure of a network, and how tree-like it is. Moreover, focusing on topological anomalies—layers whose onion spectra go against the general trends—allows us to *de novo* identify interesting subgraphs within the network. For example, without knowing what precisely we were looking for, we found backbones in networks of websites, governing

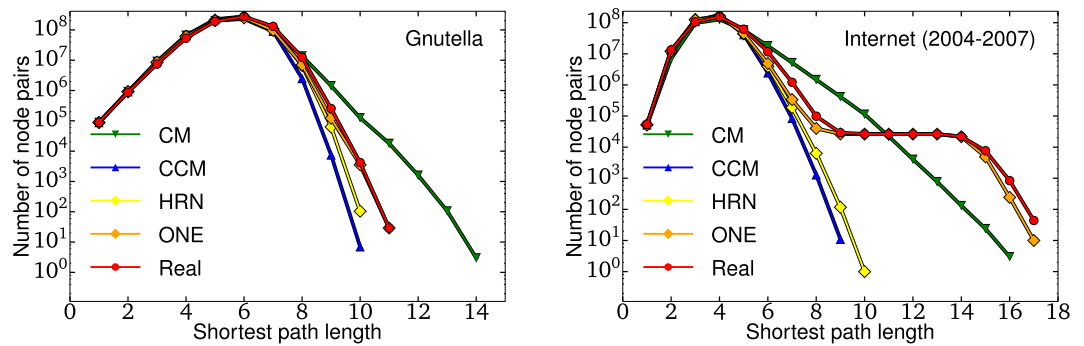


Figure 7. Distribution of the length of the shortest paths. The distribution of the length of the shortest path between all pairs of node in (left) the Gnutella peer-to-peer sharing network²² and (right) the Internet (circa 2004–2007)²⁴. The distributions of the length of shortest paths in the original networks are shown in red, and the distributions for the rewired networks according to the Configuration Model (CM), the Correlated Configuration Model (CCM), Hard-core Random Networks which approximately preserve k -core structure but not layers (HRN)²², and the ONE are shown in green, blue, yellow and orange respectively. Note that for the Internet, the CM, CCM and HRN do not even get the *shape* of the distribution correct, whereas the ONE gets not only the shape correct, but very nearly matches the exact values.

nodes connecting a set of communities, and dense communities of collaborators within a citation network. This same method could be used to find large communities within an otherwise random network, a sparse periphery to an otherwise dense network, or other interesting structures that have yet to be identified.

We then showed how all of these properties can be included in a random network ensemble by specifying connection rules that preserve a node's degree and position in the spectrum, and illustrated its potential applications by providing two examples. Future work will focus on solving the properties of that random network ensemble.

References

- Newman, M. E. J. *Networks: An Introduction* (Oxford University Press, 2010).
- Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
- Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
- Newman, M. E. J. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA* **103**, 8577–8582 (2006).
- Newman, M. E. J. Communities, modules and large-scale structure in networks. *Nature Physics* **8**, 25–31 (2012).
- Fortunato, S. Community detection in graphs. *Physics Reports* **486**, 75–174 (2010).
- Goh, K.-I., Kahng, B. & Kim, D. Universal behavior of load distribution in scale-free networks. *Phys. Rev. Lett.* **87**, 278701 (2001).
- Seidman, S. Network structure and minimum degree. *Social Networks* **5**, 269–287 (1983).
- Batagelj, V. & Zaversnik, M. An $O(m)$ algorithm for cores decomposition of networks. *arXiv preprint arXiv:cs/0310049* (2003).
- Batagelj, V. & Zaversnik, M. Generalized cores. *arXiv preprint arXiv:cs/0202039* (2002).
- Batagelj, V. & Zaversnik, M. Fast algorithms for determining (generalized) core groups in social networks. *Advances in Data Analysis and Classification* **5**, 129–145 (2011).
- Garas, A., Schweitzer, F. & Havlin, S. A k -shell decomposition method for weighted networks. *New Journal of Physics* **14**, 083030 (2012).
- Giatsidis, C., Thilikos, D. & Vazirgiannis, M. D-cores: Measuring collaboration of directed graphs based on degeneracy. *Knowledge and Information Systems* **35**, 311–343 (2013).
- Corominas-Murtra, B., Fuchs, B. & Thurner, S. Detection of the elite structure in a virtual multiplex social system by means of a generalised k -core. *PLoS ONE* **9**, e112606 (2014).
- Eidsaa, M. & Almaas, E. Investigating the relationship between k -core and s -core network decompositions. *Physica A: Statistical Mechanics and its Applications* **449**, 111–125 (2016).
- Alvarez-Hamelin, J., Dall'Asta, L., Barrat, A. & Vespignani, A. Large scale networks fingerprinting and visualization using the k -core decomposition. *Advances in Neural Information Processing Systems* 41–50 (2005).
- Brandes, U. A faster algorithm for betweenness centrality. *The Journal of Mathematical Sociology* **25**, 163–1177 (2001).
- Leskovec, J., Lang, K., Dasgupta, A. & Mahoney, M. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics* **6**, 29–123 (2009).
- Newman, M. E. J. The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA* **98**, 404–409 (2000).
- McPherson, M., Smith-Lovin, L. & Cook, J. Birds of a feather: Homophily in social networks. *Annual Review of Sociology* **27**, 415–444 (2001).
- Newman, M. E. J. & Park, J. Why social networks are different from other types of networks. *Phys. Rev. E* **68**, 036122 (2003).
- Hébert-Dufresne, L., Allard, A., Young, J.-G. & Dubé, L. J. Percolation on random networks with arbitrary k -core structure. *Phys. Rev. E* **88**, 062820 (2013).
- Anderson, R. M., May, R. M. & Anderson, B. *Infectious diseases of humans: dynamics and control* **28** (Wiley Online Library, 1992).
- Leskovec, J., Kleinberg, J. & Faloutsos, C. Graphs over time: Densification laws, shrinking diameters and possible explanations. *KDD '05 Proceedings of the eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 177–187 (2005).

Acknowledgements

The authors thank Aaron Clauset, Simon DeDeo, and Cristopher Moore for their useful comments. This work has been supported by the Santa Fe Institute, the James S. McDonnell Foundation Postdoctoral Fellowship (LHD), the Santa Fe Institute Omidyar Postdoctoral Fellowship (JAG) and the Fonds de recherche du Québec–Nature et technologies (AA).

Author Contributions

L.H.-D., J.A.G. and A.A. conceived the study. L.H.-D. and A.A. produced the results. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Hébert-Dufresne, L. *et al.* Multi-scale structure and topological anomaly detection via a new network statistic: The onion decomposition. *Sci. Rep.* **6**, 31708; doi: 10.1038/srep31708 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016