Supporting Information: Appendix

Social Success of Perfumes

Vaiva Vasiliauskaite, Tim S. Evans

Centre for Complexity Science, and Theoretical Physics Group,
Imperial College London, SW7 2AZ, U.K.

2nd April 2019

A Notation

Notation, used throughout this work is summarised in Table 1.

Notation	Description
$\mathcal{N}$	The set of notes
n, m	Notes, $n, m \in \mathcal{N}$
$\mathcal{P}$	The set of perfumes
p, q	Perfumes, $p.q \in \mathcal{P}$
$\mathcal{K}_p$	The set of notes in perfume p
$\mathcal{K}_n$	The set of perfumes containing note n
M	Perfumes with M or more ratings are considered to have reliable ratings. We chose $M = 92$ for this work. (was m)
V_p	The number of reviews ('votes') for perfume p
R_p	The average of the individual ratings for a perfume p
$\bar{R}^{(M)}$	The mean average rating R_p over all perfumes with at least m ratings
W_p	The weighted rating for perfume p
$\mathcal{G}$, $\mathbf{G}$	The perfume-note network and associated adjacency matrix
$\mathcal{H}$, $\mathbf{H}$	The enhancement network and associated adjacency matrix

Table 1: Table of notation used in this paper

There are many ways to quantify success based on customer ratings. First for each perfume p we count the number of votes V_p. Each individual reviewer i states if they 'love', 'like' or 'dislike' a perfume, a score R_{pi} which is not available to us. The web site converts this into an average rating R_p between one and five for that perfume and this is the information we have.

To perform permutation tests, we split the set of perfumes into two: one set of perfumes $\mathcal{P}^{(\mathrm{pop})}$ which contain a specified note $n^{(\mathrm{pop})}$, while the remaining perfumes without the note of interest are the subset $\mathcal{P}^{(\mathrm{reg})}$. We can then create two collections of ratings V_p, $R^{(\mathrm{pop})}$ ($R^{(\mathrm{reg})}$) with the ratings V_p of perfumes containing $n^{(\mathrm{pop})}$ (without $n^{(\mathrm{pop})}$).

$$\mathcal{P}^{(\mathrm{pop})} = \{p|(p, n^{(\mathrm{pop})}) \in \mathcal{E}\}\,, \qquad \mathcal{P}^{(\mathrm{reg})} = \{p|(p, n_{\mathrm{popular}}) \notin \mathcal{E}\}\,, \tag{A.1}$$

$$\mathcal{R}^{(\mathrm{pop})} = \{V_p|p \in \mathcal{P}^{(\mathrm{pop})}\}\,, \qquad \mathcal{R}^{(\mathrm{reg})} = \{V_p|p \in \mathcal{P}^{(\mathrm{reg})}\}\,, \tag{A.2}$$

where $\mathcal{E}$ is the set of edges in the perfume-note network $\mathcal{G}$.

To avoid the problem that a perfume with one or two perfect ratings can dominate the lists of top perfumes, we use a standard formula to rescale the ratings producing our 'weighted rating' W_p. This uses the number of ratings V_p given to a perfume and the mean rating for perfumes with a 'reasonable' average ratings equal or larger than M, which is a parameter we

have to fix. So suppose $\mathcal{P}^{(M)}$ is the set of perfumes with ratings equal to or greater than M

$$\mathcal{P}^{(M)} = \{p | R_p \geq M\} \tag{A.3}$$

The mean average rating $\bar{R}^{(M)}$ is then defined as

$$\bar{R}^{(M)} = \frac{1}{N^{(M)}} \sum_{p \in \mathcal{P}^{(M)}} V_p R_p \,, \qquad N^{(M)} = \sum_{p \in \mathcal{P}^{(M)}} V_p \tag{A.4}$$

so $N^{(M)}$ is the total number of votes given to perfumes with at least M votes each.

To avoid problems with perfumes of a few ratings, low V_p, having extreme average rating values R_p, we use then use a *weighted score* W_p. This can be defined using a Bayesian interpretation.

Suppose that all ratings are normally distributed with the same standard deviation σ but with different means, each mean corresponding to the rating for a perfume. It makes sense to choose our prior distribution to be a Normal distribution too but we are free to set the mean and standard deviation for this prior. A reasonable mean to choose is the average rating over all perfumes, $\bar{R}^{(M)}$. The standard derivation of this prior is going to be the expected error in this mean assuming we make M measurements, i.e. $\sigma_{\text{prior}} = \sigma/\sqrt{M}$. Here M is not necessarily the total number of ratings for all perfumes and instead we treat M as a parameter we can set. We have already stated that the ratings are assumed to be Normal distributed so the likelihood distribution is also a normal, with mean given by R_p, the current average rating for perfume p, and standard deviation for this average rating equal to $\sigma_p = \sigma/\sqrt{V_p}$ since we have V_p votes for perfume p. We want the expected distribution for the current average rating R_p, the posterior distribution. Combining a normally distributed prior with a normally distributed likelihood distribution can be done exactly using Bayes theorem (for example see Bailer-Jones (2017), Eq. 4.40) and we find that the posterior distribution is also a Normal distribution with mean given by the sum of means weighted by the inverses of the variances. That is the expected mean of the posterior distribution is W_p where

$$W_p = \frac{(\sigma_p)^{-2}}{(\sigma_p)^{-2} + (\sigma_{\text{prior}})^{-2}} R_p + \frac{(\sigma_{\text{prior}})^{-2}}{(\sigma_p)^{-2} + (\sigma_{\text{prior}})^{-2}} \bar{R}^{(M)} = \frac{V_p \, R_p + M \, \bar{R}^{(M)}}{V_p + M} \,. \tag{A.5}$$

This weighted score (sometimes known loosely as a 'Bayesian Rating') has often been attributed as being the basis for many online contexts such as the IMDb movie rating site, but most modern sites do not disclose their current method.

In our work we use $M = 92$ as this ensured that the mean number of votes for perfumes with at least M ratings was one standard deviation bigger than the mean number of votes for all perfumes.

B Launch date

We have a launch date for 7635 perfumes, about 72% of the perfumes in our data. The dates have been binned, mostly in decades except for the two earliest bins where there are fewer data points. The results are shown in Table 2. The majority of these perfumes in our dataset were launched relatively recently, with around 95% launched in the last twenty years. Over the last sixty years, the number of perfumes with at least one rating in our data falls off roughly exponentially with age, $\sim \exp(y/9.9)$ where y is the number of years since the perfume was launched, roughly 10% less each year we go back[1].

[1]Alternatively the form $\sim \exp(7.1 - y/8.5) + 1.0$ also gives a reasonable fit which illustrates the accuracy of these fits.

Launch date interval	Number of perfumes with ratings	Number of perfumes with ≥ 92 ratings
1533-1899	27	8
1900-1919	16	7
1920-1929	29	13
1930-1939	26	4
1940-1949	17	4
1950-1959	18	9
1960-1969	28	9
1970-1979	82	26
1980-1989	175	57
1990-1999	538	164
2000-2009	2386	609
2010-2017	4851	645
Total	8193	1555

Table 2: Popularity of perfumes by launch date. The data is sparse for very old perfumes so the time ranges are larger in that case. Otherwise, we combined perfumes into bins by decade.

C Price

We found prices for 978 of our perfumes, about 9.2% of the total and we quote results in British Pounds per 100ml.

The number of perfumes falls sharply as the increases, the number per price unit, v, falling roughly as $\sim \exp(v/70)$. Bins are in units of £50/100ml except for the most expensive perfumes where larger bins were needed. Results are shown in Table 3.

Price interval £/ 100ml	Number of perfumes with ratings	Number of perfumes with ≥ 92 ratings
0-49	516	170
50-99	238	106
100-149	98	47
150-199	35	20
200-249	19	10
250-299	13	7
300-399	14	10
400-700	5	3
Total	938	373

Table 3: Price intervals used to bin perfumes. The data is sparse for very expensive perfumes so there the price range is increased. Otherwise, we combined perfumes into bins of width £50 or 100ml.

D Network Definitions and Properties

Sets of vertices

We use two networks in our work based on perfumes and notes as nodes. A perfume will be denoted with an index p and the full set of perfumes is $\mathcal{P} = \{p_1, p_2, ..., p_{N_p}\}$, where $N_p = |\mathcal{P}|$ is the total number of perfumes in the dataset. Similarly, notes are denoted with indices n and the full set of notes is $\mathcal{N} = \{n_1, n_2, ..., n_{N_n}\}$ where $N_n = |\mathcal{N}|$. The set of notes in perfume p are denoted by $\mathcal{K}_p$ while $\mathcal{K}_n$ is the set of perfumes containing note n.

Perfume-Note Network

The *perfume-note* network, $\mathcal{G}$, is a bipartite graph. We have two types of nodes: perfumes $p \in \mathcal{P}$ and notes $n \in \mathcal{N}$. An edge is present between a note and a perfume only if that note is an ingredient of that perfume. The adjacency matrix, G_{pn} of the perfume-note network is therefore

$$G_{pn} = \begin{cases} 1 & \text{if note } n \text{ is in perfume } p \\ 0 & \text{if note } n \text{ is not in perfume } p \end{cases} . \tag{D.1}$$

In this perfume-note network $\mathcal{G}$, the neighbours of perfume p is the set $\mathcal{K}_p$ defined above, so the degree of the node associated with perfume p is $k_p = |\mathcal{K}_p| = \sum_n G_{pn}$. Likewise, in this network, the node the note n has a set $\mathcal{K}_n$ of perfumes as nearest neighbours and degree $k_n = |\mathcal{K}_n| = \sum_p G_{pn}$.

In our network, we have 10,599 perfume nodes and 990 note nodes with 89,388 edges. The network is given in Vasiliauskaite and Evans (2018). The degree distributions are shown in Fig. 1.

For the perfume nodes, we found that there are just seven perfumes that have ingredient lists of thirty or more perfumes while just 176 perfumes have twenty or more notes as their ingredients. By way of comparison, a normal distribution with the same number of perfumes having the ingredient lists of the same length as found in the data, we would expect just one perfume to have a list of length nineteen and none to have twenty or more notes in their ingredient lists. So while the distribution of the degree of perfume nodes does have a noticeable tail for large k_p, it is not too extreme. This is to be expected as there is a limit to how many ingredients you can put into a single perfume and for them all to play a significant role.

The degree distribution for the note nodes is, on the other hand, clearly fat-tailed with an equivalent normal distribution giving nodes of degrees between about 55 and 125 only. The most popular note is *Musk* used in 4768 perfumes, 44% of perfumes, the tenth most popular *Mandarin Orange* is in 1795 perfumes (17%). So it appears that like many other sets of similar objects (e.g. baby name popularity (Hahn and Bentley, 2003), dog breed popularity (Herzog et al., 2004)), there is a 'rich-get-richer' phenomena unlimited by any practical constraint leading to a fat tailed distribution in the note popularity.

Enhancement Network

Our second network representation is a directed, weighted network which we call the *enhancement network* $\mathcal{H}$.

To understand why we create we first consider perfumes with exactly the same ingredient listA couple of explanations for these come to mind. First they could be almost identical in terms of their smell. However, the second possibility is that the concentrations of individual ingredients are not at all similar. These may govern the final smell of the perfume so this pair of perfumes may not smell the same in practice. For instance, two perfumes can be composed of 3

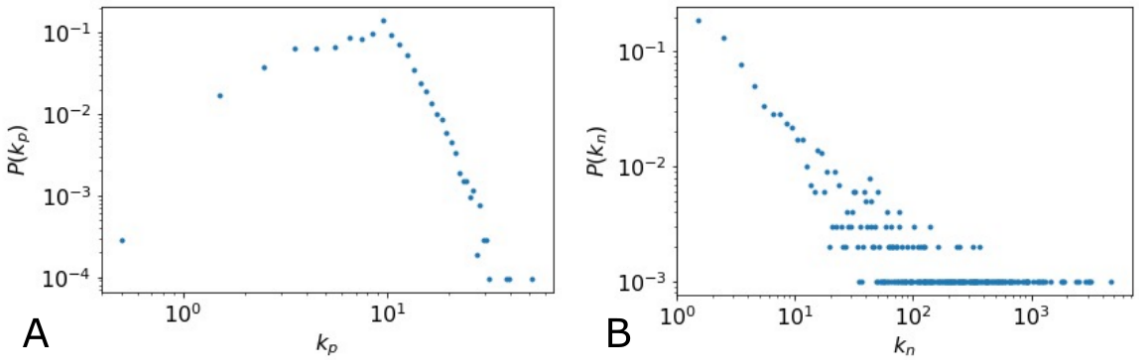

Figure 1: Degree distribution of perfumes (panel A) and notes (panel B) in the perfume-note network $\mathcal{G}$.

ingredients: musk, rose and vanilla, however ratios are, 5:2:2 and 2:5:2. The first perfume ought to smell more "musky" and the second one more "rosey". However, the detailed compositions are invariably closely guarded secrets and as we do not have precise comparisons from users in our data, we can not distinguish these two cases . There is an obvious analogy here with food recipes. The amount of chilli and the type of chilli used in a recipe can have a drastic effect on the user experience. However, like us, most food recipe network studies do not include the quantities of ingredients in their analysis.

However, we can attempt to make deductions about notes through comparisons between similar perfumes in another way. We can look at pairs of perfumes where one perfume has exactly the same ingredients as the second perfume except for the addition of **one** extra note. Our reasoning is that adding one extra note to a list of ingredients could ruin a perfume. For instance, a small drop of violet could easily overpower the entire composition, despite the amount of it in the combination being small. However, we must assume that the expert 'Nose' who created the perfume with an extra ingredient included for a good reason. So if our rating for the perfume with the extra note is higher than the perfume without the extra note, we will assume that the extra ingredient *enhances* the other notes.We will assume that the addition of an extra ingredient to a set of notes is well thought of and significantly affecting the composition overall. Our enhancement network will encode these comparisons.

Of course we still do not know if in any two perfumes differing by one note the proportions of ingredients are similar. So we will use a weighted network to capture how often we find a given enhancement. In this way we will try to use the large amount of data to build up a statistically significant picture.

Formally, we define our enhancement network $\mathcal{H}$ as follows. Each node is a note $n \in \mathcal{N}$. To define the edges, consider two perfumes that are almost identical: one perfume q has k_q notes, the set of notes $\mathcal{K}_q$, and the second perfume p has the same k_q notes plus one additional note which we call the *difference note* $n_{\rm diff}$. That is perfume p contains the notes in the set $\mathcal{K}_p = \mathcal{K}_q \cup \{n_{\rm diff}\}$. Provided the number of votes of the first perfume is smaller than that of the latter, $R_q < R_p$, the note $n_{\rm diff}$ must be enhancing the composition. We therefore draw an edge from the difference note node $n_{\rm diff}$ to the nodes representing the other notes in the two perfumes, the notes in common to both $\mathcal{K}_q$. We then add one to the weight to each edge running from the node representing $n_{\rm diff}$ to the other nodes, those in $\mathcal{K}_q$.

The adjacency matrix may be written formally as follows

$$H_{mn} \quad = \sum_{p,q\in\mathcal{P}|p\neq q} G_{pn}G_{pm}G_{qm}\delta(\mathcal{N}_p \cap \mathcal{N}_q, \mathcal{N}_q)\delta(\mathcal{N}_p \backslash \mathcal{N}_q, n)\Theta(V_p, V_q)\,. \tag{D.2}$$

Here $\delta(A, B)$ is an indicator function which is 1 if $A = B$ and 0 otherwise. The notation is a

cumbersome way of stating that perfume q has notes $\mathcal{N}_q$, containing note m, while perfume p has the same notes plus one more, note n ($n = n_{\text{diff}}$ the difference note), so that $\mathcal{N}_p = n \cup \mathcal{N}_q$. To enforce the requirement that the rating of perfume p is higher than perfume q, $V_p > V_q$, we use the indicator function $\Theta(A, B)$ which is 1 if $A > B$ and 0 otherwise. Note the edge as described above is considered to be directed from note n to note m.

Our enhancement network has 165 nodes (we ignored notes which were never involved in any enhancement), With 530 edges and a total weight of 1423. The largest weakly connected component contains 163 nodes and 529 edges (total weight 1422). The average shortest path length is 1.5 (accounting for the weights of edges).

Of course we could extend this, looking at perfumes which differ by different number of but in our opinion it is not so clear what is enhancing what in more complicated cases. Likewise our some examples differing by one note will have ratings ordered the other way round, $V_p > V_q$. We might interpret the extra note as *diminishing* the original recipe. That could be captured in a *diminishment network* $\mathcal{D}$ with adjacency matrix differing from (D.2) only in the arguments of the indicator function Θ

$$D_{mn} = \sum_{p,q \in \mathcal{P} | p \neq q} G_{pn} G_{pm} G_{qm} \delta(\mathcal{N}_p \cap \mathcal{N}_q, \mathcal{N}_q) \delta(\mathcal{N}_p \backslash \mathcal{N}_q, n) \Theta(V_q, V_p) \,. \tag{D.3}$$

It could be natural to consider these two networks together as a signed network with adjacency matrix $S_{mn} = H_{mn} - D_{mn}$.

Enhancement Network Example

In our work we apply centrality measures to our enhancement network, as discussed in Section below. In order for such measures to have meaning, we requires that paths in the enhancement network have some meaning. That is two notes linked by an edge clearly have some sort of direct relationship, but most centrality measures also use longer paths to indicate a likely if weaker relationship.

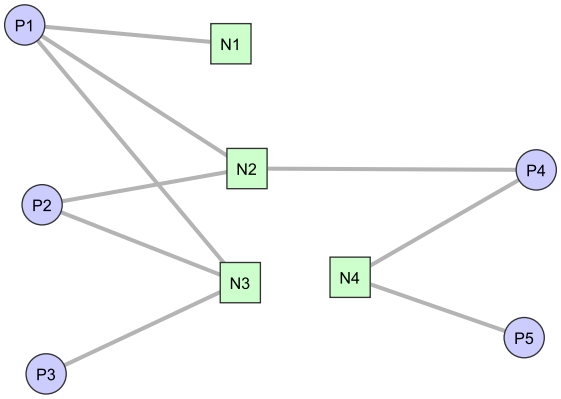

Figure 2: A very simple Perfume-Note network $\mathcal{G}_{\text{ex}}$. Note that the vertical height of the perfume notes reflects the number of votes they get; the higher up the page a perfume node is, the more votes that node has.

To see why we think paths as well as direct link may have meaning in the context of our enhancement network, we think it is helpful to look at a trivial example. Consider the perfume-note network $\mathcal{G}_{\text{ex}}$ given in Figure 2 where

- P1 (perfume 1) contains notes N1, N2 and N3,
- P2 contains notes N2, and N3,
- P3 contains notes N3,
- P4 contains notes N2 and N4,

- P5 contains note N4,
- P1 has more votes than P2, which has more votes than P3,
- P4 has more votes that P5.

Then the enhancement network $\mathcal{H}_{\text{ex}}$ derived from $\mathcal{G}_{\text{ex}}$ contains the following directed links

- from N1 to N2 (from P1 overlap with P2),
- from N1 to N3 (from P1 overlaps with P2 and P3),
- from N2 to N3 (from P2 overlap with P3),
- from N2 to N4 (from P4 overlap with P4).

This is illustrated in Figure 3.

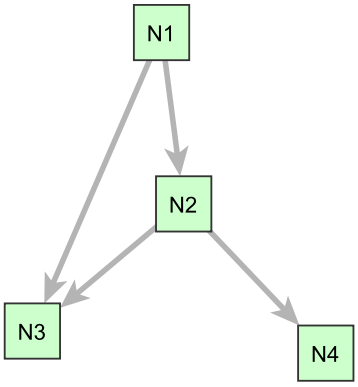

Figure 3: The enhancement network $\mathcal{H}_{\text{ex}}$ derived from the perfume-note network $\mathcal{G}_{\text{ex}}$ shown in Figure 1 where the perfume votes are as indicated in the caption of Figure 2.

Hence there is a two-step directed path $N1 \to N2 \to N3$. Here this two-step path N1 to N2 to N3 does imply some sort of relationship between the ends, between N1 and N3. After all there is also a direct link from N1 to N3 derived from the overlap of P1 and P3. This direct link will mean N3 does contribute strongly to the centrality of N1 though the indirect link will be a smaller contribution to the centrality score of N1.

Now there is also a second two-step directed path, from N1 to N2 to N4. However, in this case there is no direct edge from N1 to N4. Without a direct link, N4 will not contribute so strongly to the centrality of N1 but it be a small contribution.

The interesting point comes when we consider the importance of note N4 in any perfumes containing note N1. As with any recipe, the precise combination matters, the whole "is more than the sum of its parts". There is no direct evidence here that N1 will enhance N4 which is reflected in the lack of a direct link. This is, however, the the point of a network representation. The indirect relationship as captured by a path is some sort of indication, a suggestion or recommendation, that is a direct link between the ends of a path is likely to be a good idea. Consider a musical example, which is another nice example of a recipe. If N3 and N4 are musical notes an octave apart, or if N3 and N4 are the same musical note played by different instruments in a score, then it is a good suggestion to try notes N1 and N4 in a chord since N1 and N2 work well as musical chords.

Of course, we have no proof that indirect links are good recommendations. In our musical example, a chord N1 and N4 might be terrible because the instrument playing note N4 rather than N2 just clashes with the instrument playing note N1. The lack of a direct connection may also be a recommendation in itself, no one has made that connection for a good reason. What we are aiming to do in our analysis, as always in networks, is to look for many such

indirect suggestions. Then even if some paths represent poor suggested relationships, we hope that with many such suggestions, through the network analysis of the many possible paths in the network, the weight of good indirect relationships will reinforce each other in the way that many more poor combinations are unlikely to do.

So we assume that paths in our enhancement network can be a useful tool, just as paths rather than direct links are important in any network context. Clearly, if recipes are more than the sum of their parts, one might feel that in such contexts (music scores, perfume and food recipes etc) higher order effects (non-backtracking matrices, hypergraphs, clique overlaps etc) might provide more effective insights than simple bilateral relationships recorded in ordinary graphs. But then, that is exactly why we developed our enhancement network; a traditional network encoding higher order effects (a type of clique overlap) in the original bipartite perfume-note graph.

Incidentally, this discussion of the enhancement network and its paths also reinforces an argument why making a projection onto a network of notes alone is worthwhile. In making this one-mode projection we undoubtedly lose some information. However, the way we do the projection to create the enhancement network highlights key higher-order network effects in the original perfume-note network, higher order effects which can be drown out in standard measurements made on the original perfume-note network.

Once we have decided that paths as well as direct links in the enhancement network contain useful information then it is natural to use traditional centrality measures on the enhancement network.

Enhancement Network Centrality Measures

We used centrality measures to analyse the enhancement network. Centrality is a measure of importance of nodes, given their connections and position within a network. Newman (2010) explains in detail the majority of main centrality measures. For our purposes, we used *weighted out-degree* centrality, *out-closeness* and *reversed PageRank*, defined as follows.

Weighted out-degree s_n (strength) of a note n in the enhancement network is

$$s_n = \sum_m H_{mn} \,. \tag{D.4}$$

It is equal to the size of the multiset of notes directly enhanced by note n.

We use *out-closeness*, c_n for note n, in the form proposed by Wasserman and Faust (1994). Here we count the outward going paths as this captures chains of enhancement.

$$c_n = \frac{|\mathcal{E}_n|}{(N-1)} \frac{|\mathcal{E}_n|}{f_n} \,, \qquad f_n = \sum_{m \in \mathcal{E}_n} d_{mn} \,. \tag{D.5}$$

We define $\mathcal{E}_n$ to be the nodes which are reachable from node n (i.e. notes directly and indirectly enhanced by note n). So if $m \in \mathcal{E}_n$ there is a directed path from node n to node m. For these directed paths, we define the length of each path to be the sum of the weights along that path. Then d_{mn} is the length of the shortest path from n to m. The factor $1/(N-1)$ is an irrelevant overall constant but for completeness we note that here we have used N to be the number of nodes (notes) in the LWCC (largest weakly connected component) of the enhancement network.

A large closeness centrality score is assigned to nodes that have short paths to many nodes in the network. Thus, out-closeness centrality evaluates the universal notes potential to enhance: if a note enhances many notes, multiple times, it is assigned a large score.

We also studied *reversed PageRank*, or PageRank on the enhancement network with reversed edges. For node n it is defined to be PR_n where

$$\mathrm{PR}_n = \frac{1}{N}(1-\alpha) + \alpha \sum_m \frac{1}{k_m^{\mathrm{in}}} H_{mn} \,\mathrm{PR}_m \,, \qquad k_m^{\mathrm{in}} = \sum_m H_{mn} \,. \tag{D.6}$$

PageRank of note n counts the number notes it enhances but weights each enhanced neighbour by its importance, its PageRank. So the larger the PageRank of a note, the more important it is in terms of enhancing other notes. PageRank can be understood as a random walk where the walkers leave nodes m along incoming edges and the probability of a random walker following an incoming edge to move from m to a neighbouring node n, which is enhancing m, is proportional to the weight of the edge multiplied by α. There is also a second process, which occurs with probability $(1 - \alpha)$, where the random walker starts again from a node chosen uniformally at random from the set of nodes in the graph (a 'hyperjunp'). The probability of finding a random walker at a node in the long-time limit is the PageRank score for that node.

Several values for α were tried but we found little difference for α between 0.7 and 0.95. Therefore, we used the value $\alpha = 0.85$ which corresponds to the random walkers making on average 5.67 steps before a hyperjump compared to the the average shortest path length of 2.9.

References

C.A.L. Bailer-Jones. *Practical Bayesian Inference*. Cambridge University Press, 2017.

M. Hahn and R. Bentley. Drift as a mechanism for cultural change: an example from baby names. *Proc.R.Soc.Lon.B*, 270:S120, 2003.

H. Herzog, R. Bentley, and M. Hahn. Random drift and large shifts in popularity of dog breeds. *Proc.R.Soc.Lon B (Suppl.)*, 271:s353, 2004.

M. Newman. *Networks: An Introduction*. Oxford University Press, 2010.

V. Vasiliauskaite and T. Evans. Data for social success of perfumes, Sep 2018. doi: 10.6084/m9.figshare.7218734.

S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences)*. Cambridge University Press, 1994.