On Optimising Personal Network Size to Manage Information Flow *

Yu-En Lu
Computer Laboratory
University of Cambridge
Yu-En.Lu@cl.cam.ac.uk

Sam Roberts
Institute of Cognitive and Evolutionary Anthropology
University of Oxford
sam.Roberts@anthro.ox.ac.uk

Tammy M. K. Cheng
Institute of Biotechnology
University of Cambridge
tmkc2@cam.ac.uk

Robin Dunbar
Institute of Cognitive and Evolutionary Anthropology
University of Oxford
robin.Dunbar@anthro.ox.ac.uk

Pietro Lió
Computer Laboratory
University of Cambridge
Pietro.Lio@cl.cam.ac.uk

Jon Crowcroft
Computer Laboratory
University of Cambridge
Jon.Crowcroft@cl.cam.ac.uk

ABSTRACT

PN refer to the set of ties a specific individual has with other people. There is significant variation in the size of an individual's PN and this paper explores the effect of variation in PN size on information flow through complete social networks. We analyse degree distributions from two personal network datasets and seek to characterise PN size variations. Random matrix analysis is used to demonstrate that the specific mixture of PN sizes plays an important role in shaping the pattern of information dissemination in complete social networks. To explore this further, we conducted a series of studies on normal random graphs that represent social networks in which PN size follows a normal distribution. We demonstrate that there are three critical parameters which influence how information flows through a social network: the mean PN size, the variance in PN size and the rate at which information passes between nodes in the network. The results suggests that if the rate of information flow is increased, for example by using electronic communication rather than face-to-face communication, this could have a dramatic influence on the probability of an individual acquiring a piece of information from a person in their network.

*This research is sponsored in part by US Army Research laboratory and the UK Ministry of Defence under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the US Army Research Laboratory, the U.S. Government, the UK Ministry of Defense, or the UK Government. This work is also partly supported by the SOCIALNETS project, grant agreement number 217141, funded by the EC seventh framework programme theme FP7-ICT-2007-8.2 for Pervasive Adaptation. SR is supported by funding from the EPSRC and ESRC as part of the 'Developing Theory for Evolving Socio-Cognitive Systems' (TESS) project.

CNIKM'09, November 6, 2009, Hong Kong, China.

Categories and Subject Descriptors

G.3 [**Probability and Statistics**]: Correlation and Regression analysis; J.4 [**Computer Applications**]: Social and Behavioral Sciences

General Terms

algorithms, performance, theory

Keywords

random matrix theory, spectral graph theory, social networks

1. INTRODUCTION

In the last decade, there has been a massive increase in network research across both the social and physical sciences [1]. This has been partly due to the increased availability of electronic data on social interactions from mobile phone records and internet communication (e.g. instant messaging, email, social networking sites) that have allowed researchers to investigate social interaction on a scale not possible if relying on self-reported data from participants. In particular, research in the physical sciences has found that many different types of networks — from social networks to physical networks such as power transmission grids and the World Wide Web — have common properties such as power-law distributed degree distributions. Simulations and models of networks have produced novel findings on, for example, how information or disease spreads through networks [11].

On the other hand, there are important differences between social networks and other types of networks. In theoretical models and network simulations, the nodes are often treated as if they all have the same properties, as if a network where the nodes are people can be modelled in the same way as a network where the nodes are routers. It is clear that the "nodes" in social networks are, in reality, a heterogeneous, differentiated mass, often representing individual people with various personalities.

Thus, we start our study of social networks from the *personal network* point of view, in which the social behaviour of individuals are taken into account. There is convergent evidence to suggest that individuals have different “strategies” to build and maintain their PNs with an eye towards better location in social networks. First, all studies of personal networks have shown that there is a large variation in network size, e.g., [3]. Roberts et al. [12] demonstrated that there is a negative relationship between network size and mean emotional closeness of the network. Thus individuals tend to have either a small network of emotionally close ties or a larger network of weaker ties. This suggests both the impact of cognitive and time constraints on network size.

Second, the characteristics of personal networks have been shown to correlate with personality measures. High self-monitors — people who adjust their behaviours according to their social situation — tend to have larger networks and occupy a more strategically advantageous position in the network. Low self-monitors — people who behave consistently regardless of the social situation — have a smaller, more homogenous network of close ties.

Finally, a general “propensity to connect with others” has also been shown to be associated with having a larger network and maintaining more strategically imporant bridging ties in the workplace. Thus there is good evidence that large and small networks vary in their nature, and that at least part of the variation in network size may be due to personality factors.

1.1 Contributions

In this paper, we endeavour to start bridging the gap between the social scientists' view and that of theoreticians' by considering social networks from a *personal network* perspective. We present experiments and analyses regarding the influence of variation in personal network size on the way in which information flows through complete networks.

We analyse the impact of cognitive constraints on overall network structure in Section 3 by looking at personal networks. First, we present theoretical analysis to establish the basic spectrum distributions of general social networks. It is shown that variance of personal network size determines the entire spectrum. Through this result and spectral graph theory, to be presented in Section 2.2, this gives probabilistic bounds on the structural properties of the resulting social networks. We find that the observed distribution is such that the PN size distribution shall give rise to (global) social networks that are optimised against *relative* convergence rate of information dissemination.

2. BACKGROUND

2.1 Related Work

There has been a substantial body of work concerned with analysis of very large social networks. Here we only point to major works most directly relevant to this paper; we encourage the readers to refer to [1] for a comprehensive literature review.

Newmann [8] discussed critical conditions for random graphs of arbitrary degree distribution under which the giant-component covers the majority of the graph. Later, he investigated community structures through eigenvectors of complex networks [10], demonstrating modularity can be a real optimisation goal in any complex networks. Recently, Lewis et al. [5] have developed a dataset based on `facebook.com`, which is publicly available and which provides details of the social networks of an entire cohort of University students. The data reveal that the average size number of facebook “friends” the students have is 109, but of these only 7 are 'picture friends' — that is they have been tagged in the same picture together. This demonstrates the importance on focusing not just on the number of ties, but also on their quality. Social relationships cannot be simply reduced to a binary tie/no tie, as is the case in many network models, but are multi-faceted and vary along many different dimensions.

For the rest of this section, we present brief introductions of the theoretical and statistical tools used in our analyses. The readers can skip to the next section where appropriate.

2.2 Spectral Graph Theory

Spectral graph theory (SGT) or alternatively named algebraic graph theory characterise the relation between the eigenvalues/vectors and graphs. The main results often involve inequalities providing upper or lower bounds relating eigenvalues to graph properties. For example, the second smallest eigenvalue in the spectrum provides bounds on the connectedness of the graph and the number of zeroes give number of disconnected components in the network. Below, we give the basic formalisms from spectral graph theory so as to present our proof and simulations later.

FACT 1 (CHUNG [2]). *Let $G = (V, E)$ be a graph with vertices V and $E : V \times V$ the edges and A be its adjacency matrix as well as D be the diagonal matrix containing the degree sequences. The Laplacian matrix of G is defined as $\mathcal{L} = D^{-1/2} A D^{-1/2}$ and its eigenvalues Λ.*

The degree of connectivity is defined as

$$\phi(G) = \min_{S \subset V} \frac{E(S, \bar{S})}{\min\{|S|, |\bar{S}|\}}$$

Namely, this is the isoperimetric number of graph G. Spectral graph theory indicates that it is bounded by the

$$2\phi \geq \Lambda_2 \geq \frac{\phi^2}{2}$$

For distance between k-subgroups $\{X_1, X_2, ..., X_k : X \subset V\}$ in the system, their pairwise distance is upper-bounded by:

$$\min \mathrm{dis}(X_i, X_j) \leq \max \lceil \frac{\log \frac{\mathrm{vol}(G)}{\sqrt{vol(X_i) vol(X_j)}}}{\log \frac{\Lambda_{n-1} + \Lambda_k}{\Lambda_{n-1} - \Lambda_k}} \rceil$$

Consider a random walk matrix $P = [p_{ij}]$ over G such that

$$p_{ij} = \begin{cases} p_{ij} = \frac{1}{d_i} & \text{if } (i,j) \in E \\ 0 \end{cases}$$

There exist an equilibrium state π in which $P^t P = \pi$, provided that P is aperiodic [1]*. The rate at which a random walk to converge to π is measured in terms of* relative pointwise distance, *$\Delta(t) = \max \frac{|P^t_{ij} - \pi_i|}{\pi_i}$, is*

$$\Delta(t) \leq \exp(t(\max|1 - \Lambda_i| - 1)) \frac{\mathrm{vol}(G)}{\min_x d_x}$$

[1] For full details, see [4]

3. DEGREE MIXTURE AND SOCIAL NETWORKS

We investigate the interaction between personal network characteristics and the larger social network. We hypothesize that PN size distribution should be such that it maximises gains in certain metrics in the larger social network. Our experimental design is to synthesize social networks via random graphs that give the same PN size distribution. We note that this gives better generalisability beyond any generative model by assuming *only* link distribution (see discussion). By inspecting the spectra distribution of these random social networks, we find that PN distribution optimises relative convergence rate for random walks over social networks, as suggested by SGT results.

3.1 The Dataset

This survey followed 30 students over an 18-month period as they made the transition from school to University. Going away to University provides opportunities to form new friendships, but also places strain on existing relationships. This survey aimed to track changes in the students' social network over the course of the study, and relate this to patterns of communication.

The students completed a Social Network Questionnaire at Months 1, 9 and 18 of the study, which asked them to list the entire social network - all their living relatives, as well as all the unrelated people with whom they feel that they have a genuine personal relationship. This produced a mean network size of 51.7 (range 19-132) at Month 1. In the current analysis only the data collected at Month 1 is used, as the rest of the data is still being processed. Finally, to find additional supporting evidence to the observed distribution, we also use the published data in [12] (251 subjects) to confirm that this is a common distribution of personal network sizes.

3.2 Personal Network Size Distribution

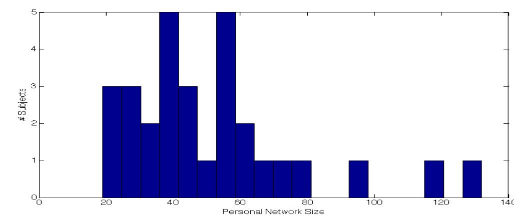
Here, we first derive the size distribution for personal networks through maximum likelyhood estimation, as shown in Figure 1. In Figure 1(a) and 1(b) we show the PN size distributions from the mobile phone dataset and another published work in [12]. Notice that the dataset in [12] consists of 251 subjects from both UK and Belgium.

However, note that both the distributions suggests a significant skew towards lower end of network size while the classic models of complex networks suggest a power-law link distribution. The existence of the skew suggests that a non-negligible number of the population have substantially fewer links than those in the right tail.

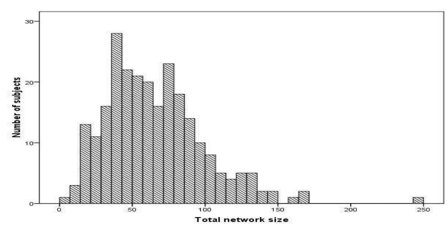
Since the degree distribution suggests a small fraction of the population having significantly larger network size, we first present a method to synthesise social networks matching the exact distributions here and then discuss our results concerning the social networks under a mixture of different personal network sizes. We will investigate whether the existence of a skewed network size distribution serve certain purposes in social networks, in the next section.

Synthesising Social Networks from Personal Networks

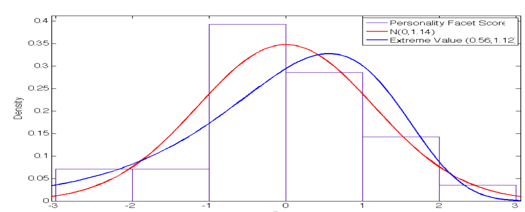
To synthesise social networks with personal networks as parameterised above, the standard Erdos-Renyi random graphs no longer suit due to the fact that its degree distribution is Binomial which is not the case in our 251 subjects and 30



(a) Measured distribution of personal network size.



(b) Measured distribution of personal network size from another dataset [12] consisting of 251 subjects. Notice the common distribution fall in the central region of the distribution. Regrettably, personality scores are not available for this dataset.



(c) Distribution fitting of the link distribution via maximum-likelyhood parameter estimation. Within 95% confidence interval the extremal value distribution fits better than Gaussian. Note that subjects falling into the left tail consist of about 15% of the samples. This corroborate with the results to be presented later in Figure 2(c).

Figure 1: Personal network size distribution.

subjects study. The problem is then how we can check the general properties of a social network that matches our observed PN size distribution.

In this paper, we propose to check *random instances* of graphs that gives the observed PN distribution. We shall focus first on how we do this and discuss *why* this can generalise to real snapshots of social networks later. We use the algorithm in [7] to synthesise social networks matching the personal network distribution discovered above. This algorithm finds a random sequence sampled from the designated distribution such that the sum of the sequence is even. Each number in this sequence denotes the degree distribution of a graph which is made complete by picking random edges between vertices that still have empty slots. This algorithm clearly terminates as the even sum sequence can be found in polynomial time with high probability and picking edges takes steps totaling the sum of the sequence. The details of the algorithm can be found in Algorithm 1.

Algorithm 1 Algorithm due to Molloy and Reed [7]

repeat
  pick $\{d_1, d_2, \cdots, d_n\}$ from distribution P_k
until $\sum d_i$ is even
for each random pair of stubs (x_i, x_j) such that $deg(x_i) < d_i \wedge deg(x_j) < d_j$ **do**
  add (x, y) into E
  $deg(x_i) \leftarrow deg(x_i) + 1$; $deg(x_j) \leftarrow deg(x_j) + 1$
end for

3.3 Spectrums of Mixture Social Networks

As shown in Figure 1, there exists a significant skew in the PN size distribution. In this section, we characterise the potential impact of the variance in PN size to how information flows through the social network as a whole. The obvious metrics to measure are global metrics such as the diameter or isoperimetric number (measuring group connectivity) of a graph. The diameter, after all, is the metric capturing the famous 6-degree of separation result by Stanley Milgram.

In theory the larger the personal networks are the more "connected" the resulting social network as a whole is. However, as we shall see below, from the personal network perspective, this claim is not straight-forward. In Theorem 1, we show that what decides the spectra for Laplacian matrices of graphs is the *variance* of link distribution, rather than sheer mean number of links. This theorem thus suggests that to maximise or minimise certain metrics, the *variance matters more than mean number of links*.

THEOREM 1. *Let* $\mathcal{L} = D^{-1/2}LD^{-1/2}$ *be the Laplacian matrix of graph* $G = (V, E)$ *where* $D = diag(d_1, d_2, ..., d_n)$, d_i *be the edge degree of vertex* i, *and* $L = [l_{ij}]$ *be such that*

$$l_{ij} = \begin{cases} \deg(i) & \text{if } i = j \\ -1 & (i,j) \in E \\ 0 & \end{cases}$$

Let Λ *be the eigenvalues of* $\mathcal{L}$ *and* σ *be the standard deviation of* $l_{1 \le i < j \le n}$. *We have*

$$\Lambda \sim 1 - D^{-1}\sqrt{n}\sigma \,\text{Semi-Circle}(2)$$

where $\text{Semi-Circle}(x; r) = \frac{1}{2\pi r^2}\int \sqrt{r^2 - x^2}dx$.

PROOF. We show this by first deriving the eigenvalue decomposition of $\mathcal{L}$ into M, the 01 adjacency matrix, and D, the diagonal matrix containing the degree sequence, from which we can gauge the distribution for each matrix. For convinience of presentation, we use $S = D^{-1/2}$.

$$\begin{aligned} \Lambda V =& S(D - M)SV \\ =& (diag(1) - SMS)\,V \\ M\,\{SV\} =& S^{-1}(diag(1) - \Lambda)S^{-1}\,\{SV\} \\ =& \lambda^M SV \end{aligned}$$

Since M is symmetric and
that diagonal matricse are commutative
Cleaning RHS and LHS, we have

$$\lambda^M = D\,(diag(1) - \Lambda)$$

Provided that $\lambda^M \sim \sqrt{n}\sigma\,\text{Semi-Circle}(2)$ (invoke Lemma 1), we arrive at the claim. □

LEMMA 1. *Let* $X_n = [x_{ij}]$ *be the adjacency graph of a* n*-node graph* $G = (V, E)$ *such that* $\max \deg(v) = c \ll n$ *and* $\lambda = \{\lambda_i\}$ *be its eigenvalues. If* $\{x_{ij}\}$ *are i.i.d. random variables such that*

$$x_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E \\ 0 & \text{otherwise} \end{cases}$$

Then we have:

$$\lambda_i \sim \sqrt{n}\sigma\,\text{Semi-Circle}(2) \tag{1}$$

where σ *is the standard deviation of* x_{ij}, $p = \Pr[x_{ij} = 1]$ *and* $\text{Semi-Circle}(r)$ *is a semi-circular distribution of radius* r.

PROOF. To apply the Semi-Circular law[2], we first define the normalised adjacency matrix $A = \frac{1}{\sigma}[X - [Ex_{ij}]]$. Let $\frac{1}{\sqrt{n}}AW = \lambda^A W$ be its eigenvalue decomposition.

Expanding the normailisation form, we have $X = \sigma A + [p]$. Applying the universality principal for shift matrices [13], we have

$$\lambda^X \sim \sqrt{n}\sigma\,\text{Semi-Circule}(2)$$

It remains to show that i) $\text{rank}([p]) = o(n)$ and ii) $\|[p]\|_2^2/2 < \infty$. Condition i) is straight-forward, note that the rank of [1] is 1. For ii), it suffices to show that $\sum_{1 \le i,j \le n} p_{ij}^2 < \infty$ as $n \to \infty$, i.e., $p \le O(n^{-1})$, which comes from the given conditions. □

3.4 Simulating Information Flow through Social Networks

Having characterised the expected behaviour of the spectra, we present simulation results for matrices with a mixture between high and low degree nodes. We first show that in the development stage during which the individuals start increasing personal network size, the social network may not benefit from the increase of links (Figure 2(a)). This is exemplified by the significant increase in variance when mixture ratio reaches 50%.

As discussed earlier in Theorem 1, the spectra is dominated by degree distribution variance. This postulates that even when subjects intend to optimise their personal network against diameter, they may find contradictory results during certain mixtures.

[2]See Section 7.1 for a introduction.

Secondly, one can also argue that subjects can only optimise using cues local to their personal network. In Figure 2(c) and 2(d), we demonstrate a "local" communication metric to that effect. This metric, $\Delta(t)$ as presented in Fact 1, aims to optimise the speed at which the social network disseminates information within the limits set by the underlying topology. One can think of this as the individual trying to make sure his/her personal network size is such that gossip spreads fastest around his/her local community.

Curiously, this simple metric reaches maximum at a certain mixture ratio and in a shape close to the observe extremal value distribution found in the network size distribution. This result indicates that, as far as disseminating information is concerned, simply increasing personal network size may not be the most direct route. This is due to the fact that increasing links means more nodes to be sent messages which in effect "widens" the necessary scope to broadcast.

To achieve maximum relative speed of convergence, the optimal strategy is in fact a certain mixture of lower degree nodes with higher degree ones, as suggested in Figure 2(c) and 2(d). Notice that these two Figures come from results based on different parameters and social network size.

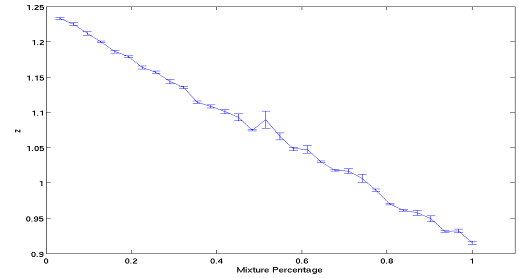
Last, but not least, we note that this metric is one of a large family of optimisation goals since its peak is determined by λ_2^M. This eigenvalue plays an important role in many similar metrics based on products of adjacency matrix and thus those can have similar behaviour to the relative convergence speed as we have seen above. In Figure 2(b), we present its simulation results over the different mixture configurations so as to ascertain its insensitivity.

3.5 Discussion

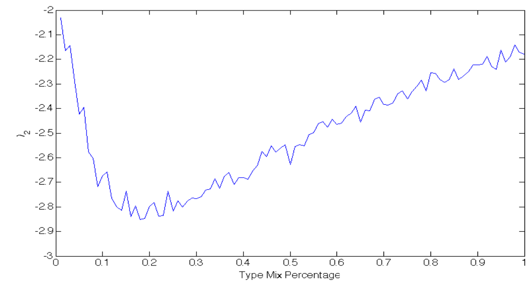
In interpreting the results of the synthesised social networks, the question arises whether the model, though it gives exact matching distribution of PN size, captures the nature of social networks. Granted that the mixture model does not take generative models such as preferential attachment [9] into account. While we admit that it *may be* the generative model for social networks, we would like to note that these experimental results are applicable to networks *with or without* a generative model. Therefore, while it is possible to construct even more realistic networks (and risk the generative model in question being wrong or involving more complications with other generative models), we resort to a very generic model that assumes nothing but a binary mixture of degree distributions. The low variance as indicated in Figure 2(c) suggests that, given the same mixture ratio, the number of graphs that offer significantly lower convergence time is very low compared to all other possibilities. Therefore, we believe that even when the effects of various generative models are taken into consideration, the general trend shown here by considering personal network size alone will remain the same.

4. PN SIZE AND INFORMATION DISTRIBUTION

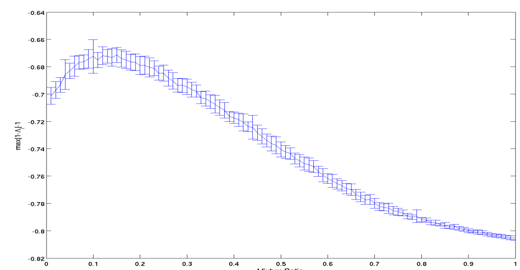
The simulation analysis demonstrated that variation in PN size influences the rate at which information flows through a global network. Nevertheless, it does not answer a more intriguing question concerning the way in which the information flow may be controlled by the PN size. Below, we present a preliminary study using very simplified parame-



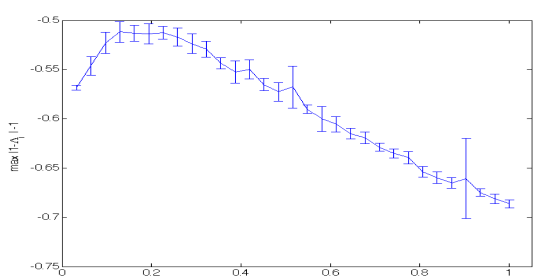
(a) The denominator of diameter upper bound, $z = \left(\log \frac{\lambda_{n-1}+\lambda_2}{\lambda_{n-1}-\lambda_2}\right)^{-1}$, steadily decreases as the proportion of nodes with high-degree exceeds that of low-degree. Notice that the variance is maximised when mixture reaches 50 : 50.



(b) λ_2^M fluctuations during distribution mixture.



(c) The exponent of random walk convergence time, $\max|1-\Lambda_i|-1$. Notice that this follows the phase transition similar to that of λ^M. Note that this reaches maxima when the mixture ratio is between 13% to 18%.



(d) The exponent of random walk convergence time, $\max|1-\Lambda_i|-1$ under mixture of $N(50, 15)$ and $N(10, 5)$. Observe that maximum is reached in similar region as above.

Figure 2: Global network spectra due to personality mixture.

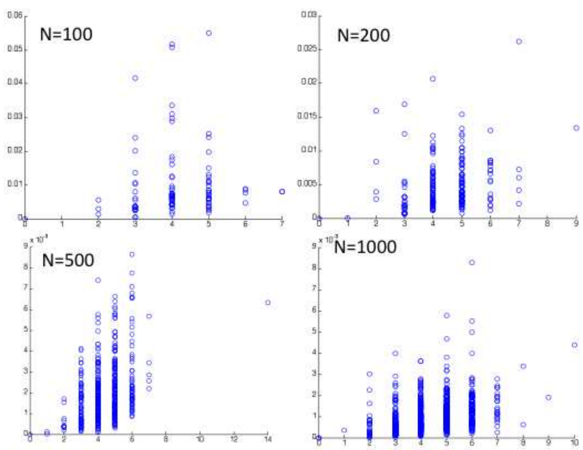


Figure 3: PN size and its influence on information flow. The x-axis indicates PN sizes as number of links; the y-axis indicates the probability of random walker to reach that node. The NRG model was generated with average PN size = 5 and standard deviation = 1.

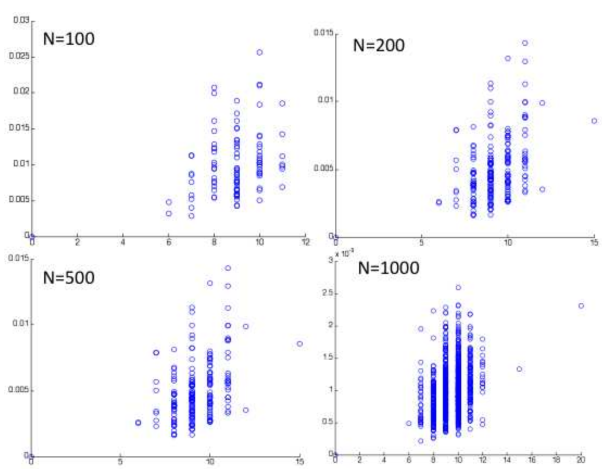


Figure 4: PN size and its influence on information flow. The x-axis indicates PN sizes as number of links; the y-axis indicates the probability of random walker to reach that node. The NRG model was generated with average PN size = 10 and standard deviation = 1.

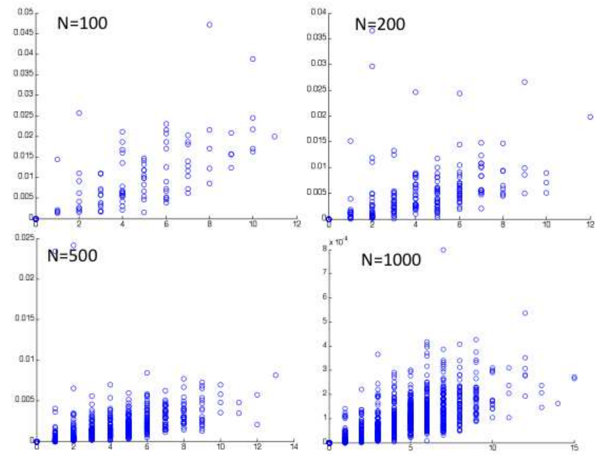


Figure 5: PN size and its influence on information flow. The x-axis indicates PN sizes as number of links; the y-axis indicates the probability of random walker to reach that node. The NRG model was generated with average PN size = 5 and standard deviation = 5.

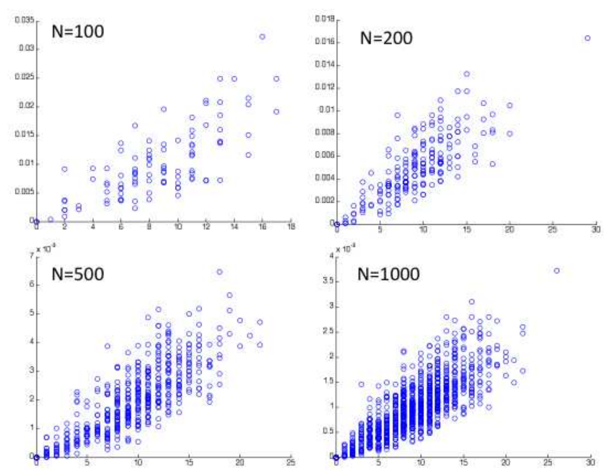


Figure 6: PN size and its influence on information flow. The x-axis indicates PN sizes as number of links; the y-axis indicates the probability of random walker to reach that node. The NRG model was generated with average PN size = 10 and standard deviation = 5.

ters to probe the basic dynamics of parameter space of PN size and each node's influence as gauged by the equilibrium probability due to the observed interaction frequency.

To further investigate the relationship between PN size and information flow, we conducted a series of studies on normal random graphs (NRG) that represent social networks in which PN size follows a normal distribution. We formulated PN size as degree of links on each node. Furthermore, we evaluated the behaviour of information transmission by estimating the *influence* of each individual, which was defined as the equilibrium probability that a certain piece of information randomly passes by (or reaches) each node. We also assigned the contact probability on each edge according to exponential distributions with lamba = 0.5.

As shown in Figure 4, when a social network has a greater PN size in general, nodes with more links tend to have a greater influence on information transmission. However, this relationship disappears when a social network contains fewer links on average – nodes that have a medium, rather than large, number of the links tend to play a more important role in distributing the information (Figure 3). Therefore, contrary to our intuition, PN size is not necessarily a good reference for identifying individuals that have a greater chance to know and pass on information. If we interpret PN size as a rough indicator for personality, i.e. nodes with larger and smaller PN sizes represent out-going and shy people respectively, the result suggests that, if a community is formed mainly by out-going people, then those have more friends have higher chance to control information flow. However, if a community does not have many out-going people, then those with neutral personality would have a higher chance of obtaining and passing on information than others. This intriguing result suggests that personality does shape the topology of the social network and thus changes the way that information flows in the network. Interestingly, such a phenomenon is independent of the total number of the people in a social network. Thus it is the personality (PN size) of the major sub-community rather than overall community size that determines how information transmits within a network.

To test if the diversity of PN size affects information transmission, we also conducted simulations on NRG models generated by different variations of PN size. As shown in Figure 5, when we increased the standard deviation of PN size (=5) in two NRG models that have average PN size = 5, the most influential nodes tend to have a small or medium number of links. In other words, when a community is less out-going and has less diversity in personality, then individuals with medium or smaller PN sizes tend to play important roles in terms of information transmission. This finding does not differ much from what we observed in the previous NRG model that has average PN size = 5 with standard deviation = 1 (Figure 3). As a comparison, we did similar simulations on NRG models consisting of more edges (Figure 6). Again, a linear relationship between the size of personal links and the influence was observed as was found in Figure 4. These results as a whole suggests that diversity of personality in a community is not as essential as which personality type was most common in terms of controlling information flow.

5. CONCLUSION AND FUTURE WORK

In this paper, we explored how variation in personal network size influenced information flow through complete social networks. Random matrix analysis demonstrated that the specific mixture of personal network sizes played an important role in shaping the speed at which information disseminated through the network. Moreover, through spectral analysis, we demonstrated that this mixture generalises to other metrics involving recursive operations of the adjacency matrix of a social network. A series of studies on normal random graphs that represent social networks, in which PN size followed a normal distribution, identified three critical parameters that influenced how information flows through a social network: the mean personal network size, the variance in personal network size and the rate at which information is passed between nodes in the network.

Electronic communication, as compared to face-to-face communication, increases the rate of information flow between individuals: it is easier to make contact with individuals electronically and thus the frequency of electronic communication tends to be higher than face-to-face contact. The results of this study suggest that this increased frequency of electronic communication may have a dramatic influence on the probablity of an individual acquiring a peice of information from a person in their network. Thus the results here have broad implications for assessing the impact of electronic communication on social relationships, as well as revealing design principles for wireless mobile networks.

Acknowledgment

YEL would like to thank Steven Hand, Andrew Moore, and Mirco Musolesi for useful suggestions and comments. All enquiries concerning the dataset used in this paper should be directed to Prof. Robin Dunbar (`Robin.Dunbar@anthro.ox.ac.uk`).

6. REFERENCES

[1] Stephen P. Borgatti, Ajay Mehra, Daniel J. Brass, and Giuseppe Labianca. Network Analysis in the Social Sciences. *Science*, 323(5916):892–895, 2009.

[2] F. Chung. *Spectral Graph Theory*. Number 92 in CBMS Regional Conference Series in Mathematics. American Mathematical Society, 1997.

[3] R. Dunbar and M. Spoors. Social networks, support cliques, and kinship. *Human Nature*, 6(3):273–290, 1995.

[4] Leonard Kleinrock. *Queueing Systems, Volume 1: Theory*. John Wiley & Sons, 1975.

[5] Kevin Lewis, Jason Kaufman, Marco Gonzalez, Andreas Wimmer, and Nicholas Christakis. Tastes, ties, and time: A new social network dataset using facebook.com. *Social Networks*, 30(4):330–342, October 2008.

[6] Madan Lal Mehta. *Random Matrices*. Academic Press, San Diego, California, second edition, 1991.

[7] Michael Molloy and Bruce Reed. A critical point for random graphs with a given degree sequence. In *Random Graphs 93: Proceedings of the sixth international seminar on Random graphs and probabilistic methods in combinatorics and computer science*, pages 161–179, New York, NY, USA, 1995. John Wiley & Sons, Inc.

[8] M. E. Newman, D. J. Watts, and S. H. Strogatz. Random graph models of social networks. *Proc Natl Acad Sci USA*, 99 Suppl 1:2566–2572, February 2002.

[9] M. E. J. Newman. Clustering and preferential attachment in growing networks. *Phys. Rev. E*, 64(2):025102, Jul 2001.

[10] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 74(3):036104, 2006.

[11] J.-P. Onnela, J. Saram£ğki, J. Hyv££nen, G. Szab££, D. Lazer, K. Kaski, J. Kert£l'sz, and A.-L. Barab£řsi. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332–7336, 2007.

[12] Sam G. Roberts, Robin I. Dunbar, Thomas V. Pollet, and Toon Kuppens. Exploring variation in active network size: Constraints and ego characteristics. *Social Networks*, February 2009.

[13] Terence Tao, Van Vu, and Manjunath Krishnapur. Random matrices: Universality of esds and the circular law, 2008.

7. APPENDIX

7.1 Random Matrix Theory

While spectral graph theory establishes the relation between spectrum and the graph, little conclusion can be made about its probabilistic behaviour. An equivalent to central limit theorem for graphs is required to estimate the general probabilistic distributions of graph spectrums. Historically, this is mostly studied in random matrix theory.

In the 1950s, it was first observed in many nuclei energy level experiments that the energy level distribution follows a distribution irrespective of the particular system state or individual particle charateristics. It was later demonstrated that the energy level (spectra) distribution depends only on the structure of the nuclei system (represented as a matrix) in question. Exact distributions for eigenvalues of matrices classes such as Gaussian and circular ensembles were derived. The conjecture was that the spectrum of matrices with i.i.d entries irrespective of distributions converge to the same empirical distribution.

It is later that Wigner derived the exact distribution — the semi-circular distribution — for random hermitian matrices with Gaussian distributed entries. Inspired by various numerical experiments later, this result was then generalised to require only finite mean and 6th moment to only finite mean and variance [3]. We state the exact form of the semi-circular law below:

FACT 2 (THE SEMI-CIRCULAR LAW). *Let $W = [w_{ij}]$ be a random real matrix of order n such that*

- *Random variables $\{w_{ij} : 1 \leq i \leq j \leq n\}$ are i.i.d.*
- *$w_{ij} = w_{ji}$ (Symmetric)*
- *$Ew_{ij} = 0$ and $var(w_{ij}) = 1$*

That is, if W is hermitian with upper-diagonal entries i.i.d with 0 mean and unit variance, then the eigenvalue distribution of $\frac{1}{\sqrt{n}}W$ converges to the empirical spectrum distribution (ESD):

$$\Pr[x] = \frac{1}{2\pi}\sqrt{4 - x^2} \qquad \text{as } n \to \infty$$

A natural question to ask is then whether this applies to systems with a prefined state. The applicability of Semi-Circular Law would be greatly reduced if it cannot give results concerning random matrices with a given starting point. The theorem below provides the necessary foundation on which we may safely ignore initial conditions of the system and analyse the convergence distribution directly. Also, it indicates that given minor purturbation, the spectrum is not easily changed. We note that this theorem generalises to non-symmetric, complex random matrices as shown in [13].

FACT 3 (UNIVERSALITY PRINCIPAL, THEOREM 1.20 [13]). *Let X_n and Y_n be $n \times n$ matrices with entries from i.i.d sequences $\{x_{i,j}\}$ and $\{y_{i,j}\}$ such that $EX = 0$ and $var(X) = 1$. For each n, let M_n be a random $n \times n$ matrix independent of X_n, Y_n such that $\Pr[\lim \frac{1}{n^2}\|M_n\|_2^2 = F] = 1$ where F is a constant. Consider the random base matrices $A_n = M_n + X_n$ and $B_n = M_n + Y_n$. Their emperical spectrum distribution* $\Pr[ESD(A_n) - ESD(B_n) \to 0] = 1$.

[3] We encourage the readers to refer to [6,13] for a review and state-of-the-art results.