Size matters: variation in personal network size, personality and effect on information transmission

Yu-En Lu* , Sam Roberts†, Pietro Lió*, Robin Dunbar†, and Jon Crowcroft*
*Computer Laboratory
University of Cambridge
{firstname.lastname}@cl.cam.ac.uk
†Institute of Cognitive and Evolutionary Anthropology
University of Oxford
{firstname.lastname}@anthro.ox.ac.uk

***Abstract*—In the last decade, there has been a massive increase in network research across both the social and physical sciences. In Physics and Mathematics, there have been extensive work on phenomenological models and generative models concerning large networks with applications to biology and social networks. In the social sciences, on the other hand, much attention has been devoted to the study of personal networks (PN) which examine the ties an individual has with others and their social characteristics and dynamics. In this paper, we seek to bridge the gap between social and mathematical sciences by exploring how variation in personal network size influences information flow through a complete network. We find that there is a significant negative correlation between a particular combination of personality traits and personal network size. A simulation modelling information flow through a complete network reveals that a mixture of small and large personal networks produces the optimal relative convergence rate at which information disseminates through the networks.**

I. 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, on how information or disease spreads through networks [2].

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.

There is convergent evidence to suggest that individuals have different "strategies" to build and maintain social networks. First, all studies of personal networks — the ties that particular individuals have with others — have shown that there is a large variation in network size, e.g., [3]. Roberts et al. [4] demonstrated that there is a negative relationship between network size and mean emotional closeness of the network. Thus individuals tend either to have 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, and also that people have preferences for different types of networks. Second, these different types of networks have been shown to correlate with personality measures. High self-monitors — that is people who adjust their behaviour 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 homogeneous network of close ties.

Further, a general "propensity to connect with others" has also been shown to be associated with having a larger network and maintaining more strategically 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.

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.

A. 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 social networks under influences of individual personality and human cognitive constraints with an 18-month longitudinal mobile-phone dataset. This dataset consists of 30 pre-university students who moved into work or higher-education in the 6th month of the study.

We first establish the role of individual personality in determining personal network sizes in Section II via non-linear principal component analysis (PCA). As the number of subjects is relatively small, we use *one principal component only* to construct personality profiles with sufficient statistical significance. Interestingly, even one dimension can exhibit strong correlation with personal network size.

We then proceed to analyse the impact of variation in personal network size on overall network structure in Section III by simulating information flow through a complete network. We find evidence that the rate of information flow through a network depends not on a static metric such as a fixed personal network size, but instead the optimal flow of information is dependent on a particular *mixture* of personal network sizes, close to that which is observed in the dataset. This corroborates the fact that the social network is not coordinated by strong global signals but is subject to various *relative optimisations* over previous states. We also present results explaining why this metric exhibits such behaviour and how it may generalise to other metrics as well.

B. Related Work

There has been much work examining large scale social networks. Here we only point major works most directly relevant to this paper; we encourage the readers to refer to [1] for a comprehensive literature review. Newmann [5] 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 [6], demonstrating modularity can be a real optimisation goal in any complex networks.

Recently, Lewis et al. [7] 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.

II. Social Links and Personality Profile

In this section, we establish how the personal network size may be influenced by personality. First we describe the dataset and related questionnaires. We then present personality profiles which give strong correlation to the dataset via Kernel PCA and discuss its implications.

A. The Dataset

This survey followed 30 students over an 18-month period as they made the transition from school to University. The sample included 15 males and 15 females. Going away to University provides opportunities to form new friendships, but also places strain on existing social 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. [8]

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. For each person in the network, we asked the students to provide information on: when they last made contact with the network member (face-to-face and non face-to-face), the mobile and landline telephone numbers of the network member, their location, how emotionally close the felt to that person (on a scale of 1-10) and the types of activities that they have done with that person over the last 6 months. In the current analysis only the data collected at Month 1 is used, as the rest of the data is still being processed.

Personality

The five-factor model of personality identifies five major components of personality - neuroticism, extraversion, openness, agreeableness and conscientiousness. These components of personality are typically identified by a factor analysis of a large number of specific personality traits (e.g. anxiety, impulsiveness, self-consciousness).

These five factors are considered "basic" for four reasons[1]. First, longitudinal studies have shown that there is relatively high temporal stability of these traits, even when the two measurements of personality are taken several years apart. Second, there is agreement across observers independently rating individuals on the personality traits. Third, the factors are consistently related to behaviour such as vocational interests, job performance, life satisfaction and academic achievement. Finally, the dimensions of personality have been found to be present in both sexes, in all races and in different cultures (see Figure 1(b) for our results confirming this). In this study, a short 50-item questionnaire was used

[1]We encourage the readers to refer to [9] for a comprehensive review

to measure the five major personality factors. This questionnaire is publicly available on the International Personality Item Pool (IPIP), and correlates highly with the longer NEO-PI-R version of the questionnaire [8].

B. Social Links and Personality

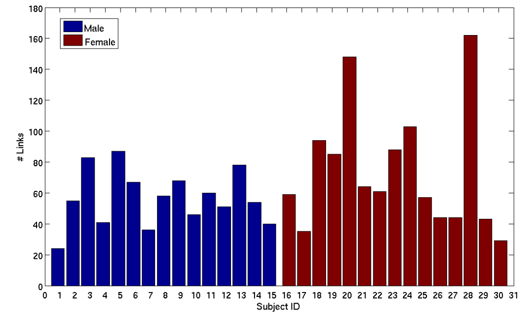
Roberts et al [10] demonstrated that there is no correlation between either extraversion or neuroticism and network size, once the effects of age are controlled for. This is contrary to the intuitive idea that at least high levels of extraversion should be associated with a larger social network. Common stereo-types often indicate that gender would play significant role. In Figure 1(a), we first show that gender and network size exhibit correlations, albeit weak.

We therefore postulate that it may be a combination of personality dimensions that explains the personal network size. The intuition is that personal network size depends on the number of people with *similar* personality instead of just individuals being out-going. The current results demonstrate that if we accept *personality profiles* — a combination of each personality dimensions — are gauged through non-linear metrics, then there exist a series of profiles that correlates strongly with personal network size.

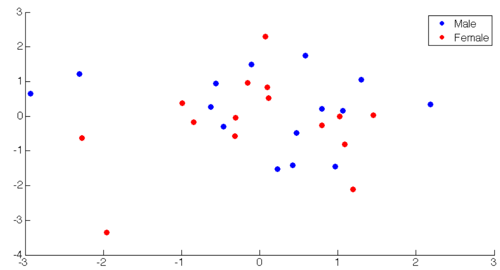
In Figure 1(c) and 1(d), we show that profiles of personality based on kernel PCA is correlated with network size. The kernel function used in the presented analysis is Gaussian 4, i.e., $K(x, y) = \exp(-\|x - y\|/4)$. We can see that the first personality profile manifested by the first kernel principal component (KPC), while offering better correlation, is weakly correlated. The fourth profile based on the fourth kernel principal component, however, yields strongest correlation with small p-value, indicating the match is of high confidence. Currently, we have evidence showing this PC is strongly correlated with the *dynamics* of social links. We would like to investigate other social characteristics and these profiles in future work.

Generalisability: We would like to note that Kernel PCAs based on Gaussian kernels with parameter σ larger than 1 all give similar correlations across each kernel principal component. Considering that this large family of Gaussian kernel PCAs (with $\sigma \geq 1$) all give good correlation, we argue that these KPCs serve as a good general model of personality profiles. Note that each Gaussian kernel is associated with a feature space of infinite dimension, known as the reproducing kernel Hilbert space (RKHS), we refer the readers to [11] for detailed information regarding its statistics and machine learning applications.

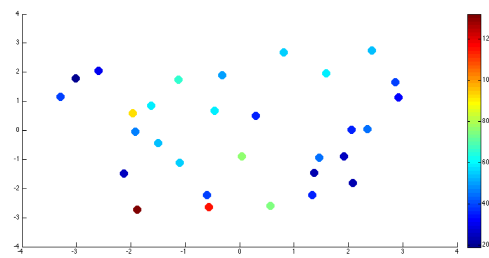
Gender and Personality: Since gender exhibits only a weak correlation with network size, it is natural to check whether at least part of the explanatory power of personality profiles is embedded within gender or vice versa. In Figure 1(b), we show that this is not the case as the spread of personality and gender is fairly even, confirming the results in psychology [9]. Therefore, the kernel components



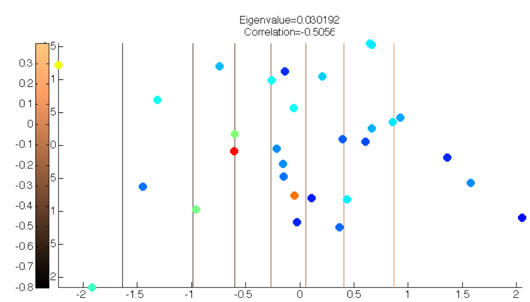
(a) Gender and network size.



(b) Gender and Personality.



(c) First and second KPCA component and network size (in color). The color bar indicates the personal network size. Notice that correlation in this case is calculated based on the First KPC only. The other KPC is here only for visualisation purposes.



(d) Fourth and fifth KPCA component and network size. The fourth KPC yields strong correlation (-0.5) with p-value $= 0.0042$. The color bar at the left indicates the value of projecting a 5D cube down to this PC. These projections forms into lines since Gaussian kernels of this parameter essentially consider points as a shell. Therefore cube shells are projected down to lines with non-linear intervals.

Figure 1. Gender, personality, and network size. Notice that the X and Y axis, where labeled, are the scores due to projections onto KPCs.

that explains the variation in personal network size owe their discriminatory power to *due to personality* rather than gender.

III. Personality Mixture and Social Networks

Having established the correlation between personality and personal network size, we would like to investigate the interaction between personal network characteristics and the larger social network. Specifically, we are interested in the impact that variation in personal networks size has on the way in which information flows through a complete social network. We design the experiment to be such that the results apply to *any random instance* of social networks matching our observed personal network characteristics.

This involves first modelling the personality and personal network size and sythesising many random social networks giving the observed quality (Section III-A). We note that this gives better generalisability across generative models by assuming *only* link distribution which we shall discuss later. By inspecting the spectra distribution of these random social networks, we find that variation in personal network size has an important influence on the rate at which information flows through a network, and that this influence would be missed if it were assumed that all individuals in the complete network have the same personal network size.

A. A Model of Personality and Links

Here, we first derive the personality distribution through maximum likelihood estimators. As discussed previously, for this dataset, it is possible to close the performance gap between KPCs and that of linear techniques by using a simple linear classifier over random projection images of personalities.

In Figure 2(a) and 2(b) we show the personal network size distributions from the mobile phone dataset and another published work in [4]. However, note that the distribution due to both approaches suggests a significant skew towards lower-end of network size with a mild two peaks at around 40 and 60, while the classic models of complex networks suggest a power-law link distribution. The existence of the skew suggests non-negligible number of the population have noticibly less links than those in the right tail.

In theory the larger the personal networks are the more "connected" the resulting social network as a whole is. Therefore, , we may expect individuals to maintain as large personal networks as they can, within cognitive constraints. However, in the experiment below, we show that the subjects *may* well be optimising their personal network such that local gossip is optimised. Before presenting our experimental results, we first provide the relation between the adjacency spectrum of the social network and its gossip convergence rate.

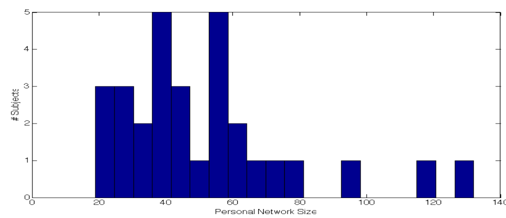
Fact 1 (Chung [12])**.** *Let $G = (V, E)$ be a graph with vertices V and $E : V \times V$ the edges and A be its adjacency matrix as well as D be the diagonal matrix containing the degree sequences. The Laplacian matrix of G is defined as $\mathcal{L} = D^{-1/2}AD^{-1/2}$ and its eigenvalues Λ.*

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

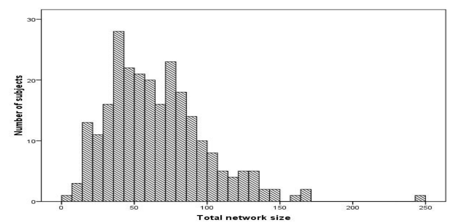
$$p_{ij} = \begin{cases} p_{ij} = \frac{1}{d_i} & \text{if } (i,j) \in E \\ 0 & \end{cases}$$

There exist an equilibrium state π in which $P^tP = \pi$, provided that P is aperiodic 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*

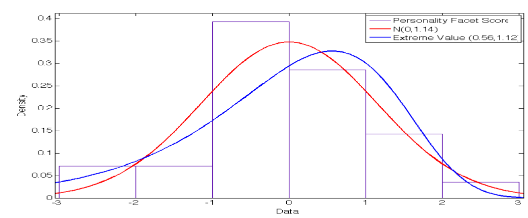
$$\Delta(t) \leq \exp(t(\max|1 - \Lambda_i| - 1))\frac{\text{vol}(G)}{\min_x d_x}$$



(a) Measured distribution of personal network size.



(b) Measured distribution of personal network size from another dataset [4] 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 Kernel PCA personality facet 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 3(c).

Figure 2. Personal network sizes and its associated personality profiles.

B. Experiment

We first show that in the development stage during which the individuals start increasing personal network size, there is a chance that the social network does not benefit from the increase of links (Figure 3(a)). As discussed earlier, the spectra is dominated by degree distribution and its variance. Therefore, it is not surprising to see variance of the measured eigenvalue fluctuate when mixture ratio reaches 50%.This postulates that even when subjects intend to optimise their personal network against diameter, they may find contradictory results during certain mixtures.

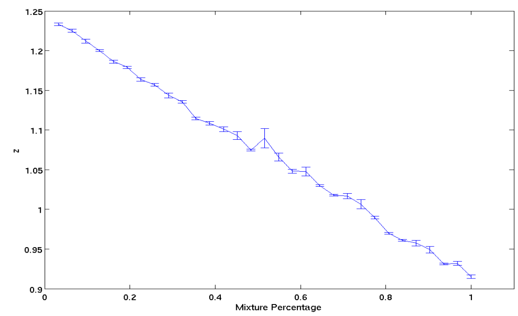
Secondly, one can also argue that subjects can only optimise using cues local to their personal network. In Figure 3(c) and 3(d), we demonstrate a "local" communication metric to that effect. This metric, $\Delta(t)$ as presented in Fact 1, aims to optimise the speed of the social network to disseminate information as far as allowed 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 and cognitive load minimal.

Curiously, this simple metric reaches maximum at a certain mixture ratio and in a shape close to the observed extremal value distribution found in the network size and personality profile distribution in Section II. This result indicates that, as far as disseminating information is concerned, simply increasing personal network size in fact decreases the rate at which information flows through the network as a whole. This is due to the fact that increasing links means more nodes to send messages to hence limits the speed at which it converges. To achieve maximum relative speed of convergence, the optimal strategy is a certain mixture of small and large personal networks, as in Figure 3(c) and 3(d). Notice that these two Figures come from results based on different parameters and social network size. Therefore, it appears that this ratio may be an mathematical optimality which should also be valid on other species and networks.

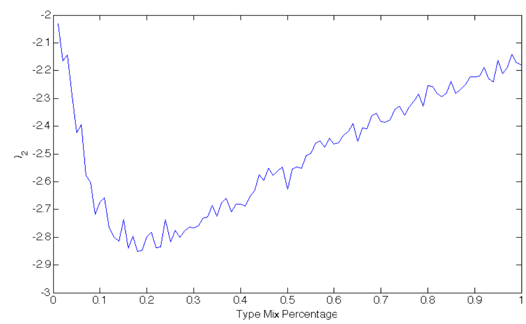
Last, but not least, we note that this metric is one of a large family of optimisation goals since its peak is determined by λ_2^M. This eigenvalue plays an important role in many similar metrics based on products of adjacency matrix and thus those can have similar behaviour to the relative convergence speed as we have seen above. In Figure 3(b), we present its simulation results over the different mixture configurations.

C. Discussion

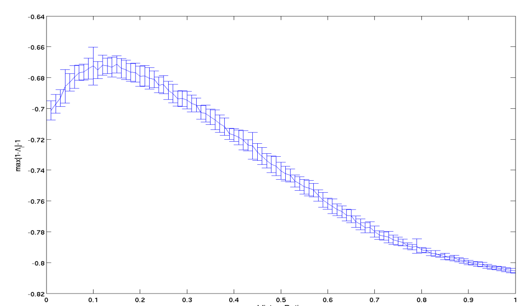
In interpreting the results of this model, the question arise whether the model captures the nature of social networks. Granted that the mixture model does not take generative models such as preferential attachment [13] into account. While we admit that it *may be the* generative model for social networks, 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 assume nothing but a binary mixture of degree



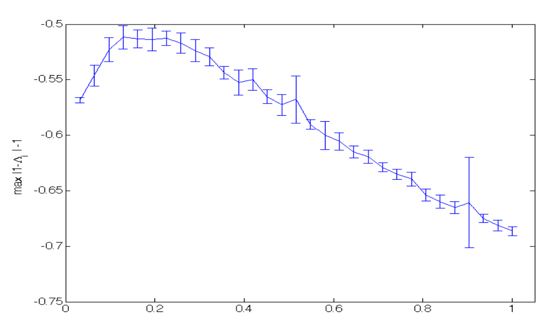
(a) The denominator of diameter upper bound, $z = \left(\log \frac{\lambda_{n-1}+\lambda_2}{\lambda_{n-1}-\lambda_2}\right)^{-1}$, steadily decreases as the proportion of nodes with high-degree exceeds that of low-degree. Notice that the variance is maximised when mixture reaches $50:50$.



(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 3. Global network spectra due to personality mixture. The social network size is set at 3000 and personal network size mixture is from $100\%N(50,15)$ to $100\%N(100,5)$.

distributions. The low variance as indicated in Figure 3(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 due to various generative models are taken into consideration, the general trend shown here by considering personal network size alone will remain the same.

IV. Conclusion and Future Work

In this paper, we present our results concerning how personality traits influence personal network size, and how this variation in personal network size influences the rate at which information flows through the social network as a whole. We found that a particular non-linear combination of personality traits is negatively correlated with personal network size, and analysed its potential linear explanations. The measured distribution is later applied to synthesize possible social networks via random graph techniques. We demonstrate that the rate at which information flows through a complete social network is in fact optimised by having a mixture of personal network sizes, rather than each individual having a large personal network size. Furthermore, this optimal mixture of personal network sizes is very close to the one observed in our dataset, and other published data.

In the future, we will use the longitdunal nature of this dataset to build dynamic social network models to exlpore how personal networks change over time.. Further, we believe the results here reveal design principles for wireless mobile networks, in that the variation in personal network size is an important variable in determining how information will flow through the complete network. In this respect, size matters.

Acknowledgment

YEL would like to thank Steven Hand, Andrew Moore, and Mirco Musolesi for useful suggestions and comments.

References

[1] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca, "Network Analysis in the Social Sciences," *Science*, vol. 323, no. 5916, pp. 892–895, 2009. [Online]. Available: http://www.sciencemag.org/cgi/content/abstract/323/5916/892

[2] J.-P. Onnela, J. Saramki, J. Hyvnen, G. Szab, D. Lazer, K. Kaski, J. Kertsz, and A.-L. Barabsi, "Structure and tie strengths in mobile communication networks," *Proceedings of the National Academy of Sciences*, vol. 104, no. 18, pp. 7332–7336, 2007. [Online]. Available: http://www.pnas.org/content/104/18/7332.abstract

[3] R. Dunbar and M. Spoors, "Social networks, support cliques, and kinship," *Human Nature*, vol. 6, no. 3, pp. 273–290, 1995. [Online]. Available: http://dx.doi.org/10.1007/BF02734142

[4] S. G. Roberts, R. I. Dunbar, T. V. Pollet, and T. Kuppens, "Exploring variation in active network size: Constraints and ego characteristics," *Social Networks*, February 2009. [Online]. Available: http://dx.doi.org/10.1016/j.socnet.2008.12.002

[5] M. E. Newman, D. J. Watts, and S. H. Strogatz, "Random graph models of social networks." *Proc Natl Acad Sci USA*, vol. 99 Suppl 1, pp. 2566–2572, February 2002. [Online]. Available: http://dx.doi.org/10.1073/pnas.012582999

[6] M. E. J. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, vol. 74, no. 3, p. 036104, 2006. [Online]. Available: http://link.aps.org/abstract/PRE/v74/e036104

[7] K. Lewis, J. Kaufman, M. Gonzalez, A. Wimmer, and N. Christakis, "Tastes, ties, and time: A new social network dataset using facebook.com," *Social Networks*, vol. 30, no. 4, pp. 330–342, October 2008. [Online]. Available: http://dx.doi.org/10.1016/j.socnet.2008.07.002

[8] T. Buchanan, J. A. Johnson, and L. R. Goldberg, "Implementing a five-factor personality inventory for use on the internet," *European Journal of Psychological Assessment*, vol. 21, no. 2, pp. 115–127, 2005. [Online]. Available: http://dx.doi.org/10.1027/1015-5759.21.2.115

[9] P. T. Costa and R. R. McCrae, "Four ways five factors are basic," *Personality and Individual Differences*, vol. 13, no. 6, pp. 653–665, 1992.

[10] S. G. Roberts and R. I. Dunbar, "Frequency of communication in active networks," *under submission*, 2009.

[11] G. Wahba, "Soft and hard classification by reproducing kernel Hilbert space methods," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 26, pp. 16 524–16 530, 2002. [Online]. Available: http://www.pnas.org/content/99/26/16524.abstract

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

[13] M. E. J. Newman, "Clustering and preferential attachment in growing networks," *Phys. Rev. E*, vol. 64, no. 2, p. 025102, Jul 2001.