A Geometric Model for On-line Social Networks*

Anthony Bonato†   *Jeannette Janssen*‡   *Paweł Prałat*§

Abstract

We study the link structure of on-line social networks (OSNs), and introduce a new model for such networks which may help infer their hidden underlying reality. In the geo-protean (GEO-P) model for OSNs nodes are identified with points in Euclidean space, and edges are stochastically generated by a mixture of the relative distance of nodes and a ranking function. With high probability, the GEO-P model generates graphs satisfying many observed properties of OSNs, such as power law degree distributions, the small world property, densification power law, and bad spectral expansion. We introduce the dimension of an OSN based on our model, and examine this new parameter using actual OSN data. We discuss how the dimension parameter of an OSN may eventually be used as a tool to group users with similar attributes using only the link structure of the network.

1. Introduction

On-line social networking sites such as Facebook, Flickr, LinkedIn, MySpace, and Twitter are examples of large-scale, complex, real-world networks, with an estimated total number of users that equals half of all Internet users [2]. We may model an OSN by a graph with nodes representing users and edges corresponding to friendship links. While OSNs gain increasing popularity among the general public, there is a parallel increase in interest in the cataloguing and modelling of their structure, function, and evolution. OSNs supply a vast and historically unprecedented record of large-scale human social interactions over time.

The availability of large-scale social network data has led to numerous studies that revealed emergent topological properties of OSNs. The next challenge is the design and rigorous analysis of models simulating these properties. Graph models were successful in simulating properties of other complex networks like the web graph (see the books [4, 7] for surveys of such models), and it is thus natural to propose models for OSNs. Few rigorous models for OSNs have been posed and analyzed, and there is no universal consensus of which properties such models should simulate. Notable recent models are those of Kumar et al. [17], which focused on simulating the component structure of OSNs, and the ILT model [5] based on transitivity properties of social networks.

Researchers are now in the enviable position of observing how OSNs evolve over time, and as such, network analysis and models of OSNs typically incorporate time as a parameter. While by no means exhaustive, some of the main observed properties of OSNs include the following.

(i) *Large-scale.* OSNs are examples of complex networks with number nodes (which we write as n) often in the millions; further, some users have disproportionately high degrees. For example, each of the nodes of Twitter corresponding to celebrities Ashton Kutcher, Ellen Degeneres, and Britney Spears have degree over four million [25].

(ii) *Small world property.* The small world property, introduced by Watts and Strogatz [26], is a central notion in the study of complex networks, and has roots in the work of the Milgram [22] on short paths of friends connecting strangers in the United States (see also [15]). The small world property demands a low diameter of $O(\log n)$, and a higher clustering coefficient than found in a binomial random graph with the same number of nodes and same average degree. Adamic et al. [1] provided an early study of an OSN at Stanford University, and found that the network has the small world property. Similar results were found in [2] which studied Cyworld, MySpace, and Orkut, and in [23] which examined data collected from Flickr, YouTube, LiveJournal, and Orkut.

*Supported by grants from MITACS, NSERC, and West Virginia University.

†Department of Mathematics, Ryerson University.

‡Department of Mathematics and Statistics, Dalhousie University.

§Department of Mathematics, West Virginia University.

In the latter study, the average distances and clustering coefficients of the OSNs were found to be lower and higher, respectively, than those of the web graph. Low diameter (of 6) and high clustering coefficient were reported in the Twitter by Java et al. [14] (see also [16]).

(iii) *Power law degree distributions.* In a graph G of order n, let N_k be the number of nodes of degree k. The degree distribution of G follows a *power law* if N_k is proportional to k^{-b}, for a fixed exponent $b > 2$. Power laws were observed over a decade ago in subgraphs sampled from the web graph, and are ubiquitous properties of complex networks (see Chapter 2 of [4]). Kumar, Novak, and Tomkins [17] studied the evolution of Flickr and Yahoo!360, and found that these networks exhibit power-law degree distributions. Golder et al. [11] discovered a power law degree distribution in the Facebook network. Power law degree distributions for both the in- and out-degree distributions were documented in Flickr, YouTube, LiveJournal, and Orkut [23], as well as in Twitter [14].

(iv) *Shrinking distances.* Kumar et al. [17] reported that in Flickr and Yahoo!360 the diameter actually decreases over time. Similar results were reported for Cyworld in [2]. Well-known models for complex networks such as preferential attachment or copying models have logarithmically growing diameters with time. Various models (see [18, 19]) were proposed simulating power law degree distributions and decreasing distances.

(v) *Bad spectral expansion.* Social networks often organize into separate clusters in which the intra-cluster links are significantly higher than the number of inter-cluster links. In particular, social networks contain communities (characteristic of social organization), where tightly knit groups correspond to the clusters [24]. As a result, it is reported in [9] that social networks, unlike other complex networks, possess bad spectral expansion properties realized by small gaps between the first and second eigenvalues of their adjacency matrices.

Our main contributions in the present work are to provide a model—the geo-protean (GEO-P) model—which simulates all five properties above (see Section 3), and to suggest a reverse engineering approach to OSNs. Given only the link structure of OSNs, we ask whether it is possible to infer the hidden reality of such networks. Can we group users with similar attributes from only the link structure? For instance, a reasonable assumption is that out of the millions of users on a typical OSN, if we could assign the users various attributes such as age, sex, religion, geography, and so on, then we should be able to identify individuals or at least small sets of users by their set of attributes. Thus, if we can infer a set of identifying attributes for each node from the link structure, then we can use this information to recognize communities and understand connections between users.

Characterizing users by a set of attributes leads naturally to a vector-based or geometric approach to OSNs. In geometric graph models, nodes are identified with points in a metric space, and edges are introduced by probabilistic rules that depend on the proximity of the nodes in the space. We envision OSNs as embedded in a *social space*, whose dimensions quantify user traits such as interests or geography; for instance, nodes representing users from the same city or in the same profession would likely be closer in social space. A first step in this direction was given in [20], which introduced a rank-based geometric model for social networks. Such an approach was taken in the SPA geometric model for the web graph; see [3]. In [13] it is shown how a theoretical analysis of the SPA model leads to a highly accurate measure for the estimated spatial distance between nodes, based on the number of common neighbours.

The geo-protean model incorporates a geometric view of OSNs, and also exploits ranking to determine the link structure. Higher ranked nodes are more likely to receive links. A formal description of the model is given in Section 2. Results on the model are summarized in Section 3. We present a novel approach to OSNs by assigning them a dimension; see the formula (4). Given certain OSN statistics (order, power law exponent, average degree, and diameter), we can assign each OSN a dimension based on our model. In a certain sense, the dimension of an OSN is the least integer m such that we can accurately embed the OSN in m-dimensional Euclidean space. A suggestive (although unproven) interpretation of the dimension is that it equals the least number of attributes needed to identify individuals or small sets of users. A sketch of proofs of our results are presented in Section 4. We conclude with a discussion of our findings, and present an outline of future work.

2. The GEO-P Model for OSNs

We now present our model for OSNs, which is based on both the notions of embedding the nodes in a metric space (geometric), and a link probability based on a ranking of the nodes (protean). We identify the users of an OSN with points in m-dimensional Euclidean space. Each node has a region of influence, and nodes may be joined with a certain probability if they land within each others region of influence. Nodes are ranked by their popularity from 1 to n, where n is the number of nodes, and 1 is the highest ranked node. Nodes that are ranked higher have larger regions of influence, and so are more likely to acquire links over time. For simplicity, we consider only undirected graphs. The number of nodes n is fixed but the model is dynamic: at each time-step, a node is born and one dies. A static number of nodes is more representative of the reality of OSNs, as the number of users in an OSN would typically have a maximum (an absolute maximum arises from roughly the number of users on the internet, not counting multiple accounts). For a discussion of ranking models for complex networks, see [10, 12, 21].

We now formally define the GEO-P model. The model produces a sequence $(G_t : t \geq 0)$ of undirected graphs on n nodes, where t denotes time. We write $G_t = (V_t, E_t)$. There are four parameters: the *attachment strength* $\alpha \in (0, 1)$, the *density parameter* $\beta \in (0, 1-\alpha)$, the *dimension* $m \in \mathbb{N}$, and the *link probability* $p \in (0, 1]$. Each node $v \in V_t$ has rank $r(v, t) \in [n]$ (we use $[n]$ to denote the set $\{1, 2, \ldots, n\}$). The rank function $r(\cdot, t) : V_t \to [n]$ is a bijection for all t, so every node has a unique rank. The highest ranked node has rank equal to 1; the lowest ranked node has rank n. The initialization and update of the ranking is done by *random initial rank*. In particular, the node added at time t obtains an initial rank R_t which is randomly chosen from $[n]$ according to a prescribed distribution. Ranks of all nodes are adjusted accordingly. Formally, for each $v \in V_{t-1}$,

$$r(v, t) = r(v, t - 1) + \delta - \gamma,$$

where $\delta = 1$ if $r(v, t - 1) > R_t$ and 0 otherwise, and $\gamma = 1$ if the rank of the node deleted in step t is smaller than $r(v, t - 1)$, and 0 otherwise.

Let S be the unit hypercube in $\mathbb{R}^m$, with the torus metric $d(\cdot, \cdot)$ derived from the L_∞ metric. In particular, for any two points x and y in $\mathbb{R}^m$,

$$d(x, y) = \min\{||x - y + u||_\infty : u \in \{-1, 0, 1\}^m\}.$$

The torus metric thus "wraps around" the boundaries of the unit cube, so every point in S is equivalent. The torus metric is chosen so that there are no boundary effects, and altering the metric will not significantly affect the main results.

To initialize the model, let $G_0 = (V_0, E_0)$ be any graph on n nodes that are chosen from S. We define the *influence region* of node v at time $t \geq 0$, written $R(v, t)$, to be the ball around v with volume

$$|R(v, t)| = r(v, t)^{-\alpha} n^{-\beta} .$$

For $t \geq 1$, we form G_t from G_{t-1} according to the following rules.

(i) Add a new node v that is chosen *uniformly at random* from S. Next, independently, for each node $u \in V_{t-1}$ such that $v \in R(u, t - 1)$, an edge vu is created with probability p. Note that the probability that u receives an edge is proportional to $p\, r(u, t - 1)^{-\alpha}$. The negative exponent guarantees that nodes with higher ranks ($r(u, t - 1)$ close to 1) are more likely to receive new edges than lower ranks.

(ii) Choose uniformly at random a node $u \in V_{t-1}$, delete u and all edges incident to u.

(iii) Update the ranking function $r(\cdot, t) : V_t \to [n]$.

Since the process is an ergodic Markov chain, it will converge to a stationary distribution. The random graph corresponding to this distribution with given parameters α, β, m, p is called the *geo-protean* (or *GEO-P* model) graph, and is written GEO-P(α, β, m, p). See Figure 1 for a simulation of the model in the unit square.

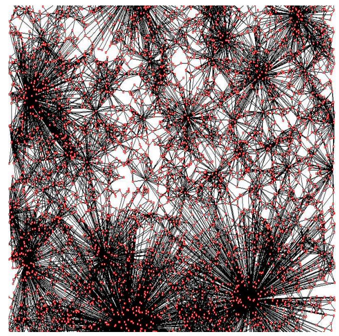

Figure 1: A simulation of the GEO-P model, with $n = 5,000$, $\alpha = 0.7$, $\beta = 0.15$, $m = 2$, and $p = 0.9$.

3. Results and Dimension

We now state the main theoretical results on the geo-protean model. The model generates with high probability graphs satisfying each of the properties (i) to (v) in the introduction. Proofs are sketched (or omitted due to lack of space) in Section 4; complete proofs will appear in the full version. Throughout, we will use the stronger notion of *wep* in favour of the more commonly used *aas*, since it simplifies some of our proofs. We say that an event holds *with extreme probability* (*wep*), if it holds with probability at least $1 - \exp(-\Theta(\log^2 n))$ as $n \to \infty$. Thus, if we consider a polynomial number of events that each holds *wep*, then *wep* all events hold.

Let $N_k = N_k(n, p, \alpha, \beta)$ denote the number of nodes of degree k, and $N_{\geq k} = \sum_{l \geq k} N_l$. The following theorem demonstrates that the geo-protean model generates power law graphs with exponent

$$b = 1 + 1/\alpha. \tag{1}$$

Note that the variables $N_{\geq k}$ represent the cumulative degree distribution, so the degree distribution has power law exponent $1/\alpha$.

Theorem 3.1. *Let $\alpha \in (0,1)$, $\beta \in (0,1-\alpha)$, $m \in \mathbb{N}$, $p \in (0,1]$, and*

$$n^{1-\alpha-\beta}\log^{1/2} n \le k \le n^{1-\alpha/2-\beta}\log^{-2\alpha-1} n.$$

Then wep *GEO-P(α,β,m,p) satisfies*

$$N_{\ge k} = \left(1+O(\log^{-1/3} n)\right)\frac{\alpha}{\alpha+1}p^{1/\alpha}n^{(1-\beta)/\alpha}k^{-1/\alpha}.$$

For a graph G of order n, define the *average degree of* G by $d = \frac{2|E|}{n}$. Our next results shows that geo-protean graphs are dense.

Theorem 3.2. Wep *the average degree of GEO-P(α,β,m,p) is*

$$d \quad = \quad (1+o(1))\frac{p}{1-\alpha}n^{1-\alpha-\beta}. \tag{2}$$

Note that the average degree tends to infinity with n; that is, the model generates graphs satisfying a *densification power law*. In [18], densification power laws were reported in several real-world networks such as the physics citation graph and the internet graph at the level of autonomous systems.

Our next result describes the diameter of graphs sampled from the GEO-P model. While the diameter is not shrinking, it can be made constant by allowing the dimension to grow as a function of n.

Theorem 3.3. *Let $\alpha \in (0,1)$, $\beta \in (0,1-\alpha)$, $m \in \mathbb{N}$, and $p \in (0,1]$. Then* wep *the diameter of GEO-P(α,β,m,p) is*

$$O(n^{\frac{\beta}{(1-\alpha)m}}\log^{\frac{2\alpha}{(1-\alpha)m}} n). \tag{3}$$

We note that in a geometric model where regions of influence have constant volume and possessing the same average degree as the geo-protean model, the diameter is $\Theta(n^{\frac{\alpha+\beta}{m}})$. This is a larger diameter than in the GEO-P model. If $m = C\log n$, for some constant $C > 0$, then *wep* we obtain a diameter bounded by a constant. We can also prove (and will be presented in the full version) that *wep* the GEO-P model generates graph with constant clustering coefficient.

The normalized Laplacian of a graph, introduced by Chung [6], relates to important graph properties. Let A denote the adjacency matrix and D denote the diagonal degree matrix of a graph G. Then the normalized Laplacian of G is $\mathcal{L} = I - D^{-1/2}AD^{-1/2}$. Let $0 = \lambda_0 \le \lambda_1 \le \cdots \le \lambda_{n-1} \le 2$ denote the eigenvalues of $\mathcal{L}$. The *spectral gap* of the normalized Laplacian is

$$\lambda = \max\{|\lambda_1 - 1|, |\lambda_{n-1} - 1|\}.$$

A small spectral gap is an indication of bad expansion properties, which are characteristic of OSNs (see property (v) in the introduction).

The following theorem suggests a significantly smaller spectral difference between graphs generated by our model than in random graphs.

Theorem 3.4. *Let $\alpha \in (0,1)$, $\beta \in (0,1-\alpha)$, $m \in \mathbb{N}$, and $p \in (0,1]$. Let $\lambda(n)$ be the spectral gap of the normalized Laplacian of GEO-P(α,β,m,p). Then* wep

(i) If $m = m(n) = o(\log n)$, then $\lambda(n) = 1 + o(1)$.

(ii) If $m = m(n) = C\log n$ for some $C > 0$, then

$$\lambda(n) \ge 1 - \exp\left(-\frac{\alpha+\beta}{C}\right).$$

Theorem 3.4 represents a drastic departure from the good expansion found in random graphs, where $\lambda = o(1)$ [6, 7].

3.1. Dimension of OSNs

Given an OSN, we describe how we may find the corresponding dimension parameter m if we assume the GEO-P model. In particular, if we know the order n, power law exponent b, average degree d, and diameter D of an OSN, then we can calculate m using the formulas (1), (2), and (3). The formula for m then (ignoring constants in the $O(\cdot)$ and $\Omega(\cdot)$ notation) becomes

$$m = \frac{\log\left(\frac{n}{2d^{\left(\frac{b-1}{b-2}\right)}}\right)}{\log D} \tag{4}$$

Note that (4) suggests a logarithmic growth for the dimension m depending on n.

The parameters b, d, and D have been determined for samples from OSNs in various studies such as [2, 23, 14]. The following chart summarizes this data and gives the dimension for each network, with citations (next to the name of the network) where the data was taken. We round m up to the nearest integer. Owing to the large-scale character of these networks, our assumption is that the parameters b, d, D found in those samples approximate the ones in the entire network. We note that the estimates of the number of users n for Flickr and Twitter come from Wikipedia [27], those from YouTube comes from their website [28], and those of Cyworld are from [8]. When the data consisted of directed graphs, we took b to be the power law exponent for the in-degree distribution. As noted in [2], the power law exponent of $b = 5$ for Cyworld holds only for users whose degree is at most approximately 100.

Parameter	**OSN**	
	Cyworld [2]	Flickr [23]
n	1.2×10^7	3.2×10^7
b	5	2.78
d	31.6	12.24
D	6	27
m	7	4

Parameter	**OSN**	
	Twitter [14]	YouTube [23]
n	4.55×10^7	3×10^8
b	2.4	2.99
d	18.86	4.29
D	6	21
m	4	6

4. Proofs of Results

Owing to space constraints, we only give sketches of the proofs of the degree distribution and diameter results of Section 3. Complete proofs of all the results will appear in the full version of the paper. The following theorem (whose proof is omitted) shows how the degree of a given node depends on the *age rank*; that is, the bijection $a(\cdot, t) : V_t \to [n]$ where nodes are ranked by age (the oldest node has rank equal to 1, and the youngest one has rank n).

Theorem 4.1. *Let $\alpha \in (0,1)$, $\beta \in (0, 1-\alpha)$, $m \in \mathbb{N}$, $p \in (0,1]$, $i = i(n) \in [n]$. Let v_i be the node in GEO-P(α, β, m, p) whose age rank at time L equals $a(v_i, L) = i$, and let R_i be the initial rank of v_i.*

If $R_i \geq \sqrt{n}\log^2 n$, then wep

$$\deg(v_i,L)=(1+O(\log^{-1/2} n))p\Big(\tfrac{i}{(1-\alpha)n}+\big(\tfrac{R_i}{n}\big)^{-\alpha}\tfrac{n-i}{n}\Big)n^{1-\alpha-\beta}.$$

Otherwise, that is, if $R_i < \sqrt{n}\log^2 n$, wep

$$\deg(v_i,L) \geq (1+O(\log^{-1/2} n)) \quad p\big(\tfrac{i}{(1-\alpha)n}+n^{\alpha/2}\log^{-2\alpha} n\,\tfrac{n-i}{n}\big) \times n^{1-\alpha-\beta}.$$

Theorem 4.1 implies that *wep* the minimum degree is $(1+o(1))pn^{1-\alpha-\beta}$. Moreover, the average degree is

$$\begin{aligned} d &= (1+o(1))\frac{2}{n}\sum_{i=1}^{n}\frac{p}{1-\alpha}\,\frac{i-1}{n-1}n^{1-\alpha-\beta} \\ &= (1+o(1))\frac{p}{1-\alpha}n^{1-\alpha-\beta}. \end{aligned}$$

Hence, the proof of Theorem 3.2 follows.

Proof of Theorem 3.1. By Theorem 4.1 it follows that *wep* each node v_i that has the initial rank $R_i \geq \sqrt{n}\log^2 n$ such that

$$\frac{R_i}{n} \geq \big(1+\log^{-1/3} n\big)\left(pn^{1-\alpha-\beta}\frac{n-i}{n}k^{-1}\right)^{1/\alpha}$$

has fewer than k neighbours, and each node v_i for which

$$\frac{R_i}{n} \leq \big(1-\log^{-1/3} n\big)\left(pn^{1-\alpha-\beta}\frac{n-i}{n}k^{-1}\right)^{1/\alpha}$$

has more than k neighbours.

Let i_0 be the largest value of i such that

$$\left(pn^{1-\alpha-\beta}\frac{n-i}{n}k^{-1}\right)^{1/\alpha} \geq \frac{2\log^2 n}{\sqrt{n}}.$$

(Note that $i_0 = n - O(n/\log n)$, since $k \leq n^{1-\alpha/2-\beta}\log^{-2\alpha-1} n$.) Thus,

$$\begin{aligned} \mathbb{E}N_{\geq k} &= \textstyle\sum_{i=1}^{i_0}\big(1+O(\log^{-1/3} n)\big)\big(pn^{1-\alpha-\beta}\frac{n-i}{n}k^{-1}\big)^{1/\alpha} \\ &\quad +O\Big(\textstyle\sum_{i=i_0+1}^{n}\frac{\log^2 n}{\sqrt{n}}\Big) \\ &= \big(1+O(\log^{-1/3} n)\big)\big(pn^{-\alpha-\beta}k^{-1}\big)^{1/\alpha}\textstyle\sum_{i=1}^{n}(n-i)^{1/\alpha} \\ &= \big(1+O(\log^{-1/3} n)\big)\big(pn^{-\alpha-\beta}k^{-1}\big)^{1/\alpha}\frac{n^{1+1/\alpha}}{1+1/\alpha} \\ &= \big(1+O(\log^{-1/3} n)\big)\frac{\alpha}{\alpha+1}p^{1/\alpha}n^{(1-\beta)/\alpha}k^{-1/\alpha}. \end{aligned}$$

and the assertion follows from the Chernoff bound, since $k \leq n^{1-\alpha/2-\beta}\log^{-2\alpha-1} n$ and so $\mathbb{E}N_{\geq k} = \Omega(\sqrt{n}\log^{2+1/\alpha} n)$. □

We now consider the diameter of graphs generated by the GEO-P model.

Proof of Theorem 3.3. We partition the hypercube into $1/A$ hypercubes, each of volume A (with A to be determined later). Consider nodes with initial rank at most R and age at most $n/2$ (again, R to be determined later). In each small hypercube, we would like to have $\log^2 n$ such nodes whose initial influence region contains the whole hypercube in which the node is located, and also all neighbouring hypercubes.

It can be shown that *wep* this is the case and remains so to the end of the process, if the initial influence region is slightly larger than $4^m A$; say, at least $5^m A$. Therefore, we obtain that

$$R^{-\alpha}n^{-\beta} = 5^m A. \tag{5}$$

It follows from the Chernoff bound that *wep* every small hypercube contains $\log^2 n$ such nodes, provided that the expected number is at least, say, $2\log^2 n$. Hence, we require that

$$\frac{n}{2}A\frac{R}{n} = 2\log^2 n. \tag{6}$$

Combining (5) and (6) we obtain that the number of hypercubes is equal to

$$\frac{1}{A} = 5^{\frac{m+\alpha}{1-\alpha}}n^{\frac{\beta}{1-\alpha}}\log^{\frac{2\alpha}{1-\alpha}} n.$$

Now, since *wep* there are $\log^2 n$ nodes in each hypercube to choose from, *wep* we can select exactly one node from each hypercube so that each node is adjacent to the chosen nodes from all neighbouring hypercubes (the younger node falls into the region of influence of the older neighbours, and creates an edge with probability p). Let us call this subgraph the *backbone*. It is clear that the diameter of the backbone is

$$\left(\frac{1}{A}\right)^{1/m} = O(n^{\frac{\beta}{(1-\alpha)m}} \log^{\frac{2\alpha}{(1-\alpha)m}} n)$$

To finish the proof, we will show that *wep* any node v that is not in the backbone is within distance two from some node in the backbone. Since *wep* the minimum degree is $\Omega(n^{1-\alpha-\beta})$, *wep* $\Omega(n^{1-\alpha-\beta})$ neighbours of v have age rank at least $n/2$. Since each such neighbour falls into the region of influence of some node in the backbone, *wep* at least one neighbour of v must be connected the backbone. □

5. Conclusion and Discussion

We introduced the geo-protean (GEO-P) geometric model for OSNs, and showed that with high probability, the model generates graphs satisfying each of the properties (i) to (v) in the introduction. We introduce the dimension of an OSN based on our model, and examine this new parameter using actual OSN data. We observed that the dimension of various OSNs ranges from four to 7. It may therefore, be possible to group users via a relatively small number of attributes, although this remains unproven.

The ideas of using geometry and dimension to explore OSNs is a novel one, and deserves to be more thoroughly investigated. It would be interesting to explore how best to fit OSNs into Euclidean space, with an emphasis on the accuracy of the embedding. Our hope is that a geometric structure exhibited by social networks may be used to further explore properties of the network. Can a geometric embedding of OSNs help determine their community structure? Another direction is to examine the edge-length distribution in the GEO-P model, and then apply that to actual OSN to classify “long” links in social networks.

References

[1] L.A. Adamic, O. Buyukkokten, E. Adar, A social network caught in the web, *First Monday* **8** (2003).

[2] Y. Ahn, S. Han, H. Kwak, S. Moon, H. Jeong, Analysis of topological characteristics of huge on-line social networking services, In: *Proceedings of the 16th International Conference on World Wide Web*, 2007.

[3] W. Aiello, A. Bonato, C. Cooper, J. Janssen, P. Prałat, A spatial web graph model with local influence regions, *Internet Mathematics* **5** (2009), 175–196.

[4] A. Bonato, *A Course on the Web Graph*, American Mathematical Society Graduate Studies Series in Mathematics, Providence, Rhode Island, 2008.

[5] A. Bonato, N. Hadi, P. Horn, P. Prałat, C. Wang, Models of on-line social networks, accepted to *Internet Mathematics*, 2010.

[6] F.R.K. Chung, *Spectral Graph Theory*, American Mathematical Society, Providence, Rhode Island, 1997.

[7] F.R.K. Chung, L. Lu, *Complex Graphs and Networks*, American Mathematical Society, U.S.A., 2004.

[8] Cyworld: South Korea’s leading social network. Accessed February 15, 2010. `http://www.squidoo.com/cyworld`

[9] E. Estrada, Spectral scaling and good expansion properties in complex networks, *Europhys. Lett.* **73** (2006) 649–655.

[10] S. Fortunato, A. Flammini, F. Menczer, Scale-free network growth by ranking, *Phys. Rev. Lett.* **96**(21): 218701 (2006).

[11] S. Golder, D. Wilkinson, B. Huberman, Rhythms of social interaction: messaging within a massive online network, In: *3rd International Conference on Communities and Technologies*, 2007.

[12] J. Janssen, P. Prałat, Protean graphs with a variety of ranking schemes, *Theoretical Computer Science* **410** (2009), 5491–5504.

[13] J. Janssen, P. Prałat, R. Wilson, Estimating node similarity from co-citation in a spatial graph model, In: *Proceedings of 2010 ACM Symposium on Applied Computing–Special Track on Self-organizing Complex Systems*, 5pp.

[14] A. Java, X. Song, T. Finin, B. Tseng, Why we twitter: understanding microblogging usage and communities, In: *Proceedings of the Joint 9th WEBKDD and 1st SNA-KDD Workshop 2007*, 2007.

[15] J. Kleinberg, The small-world phenomenon: An algorithmic perspective, In: *Proceedings of the 32nd ACM Symposium on Theory of Computing*, 2000.

[16] B. Krishnamurthy, P. Gill, M. Arlitt, A few chirps about Twitter, In: *Proceedings of The First ACM SIGCOMM Workshop on Online Social Networks*, 2008.

[17] R. Kumar, J. Novak, A. Tomkins, Structure and evolution of on-line social networks, In: *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2006.

[18] J. Leskovec, J. Kleinberg, C. Faloutsos, Graphs over time: densification Laws, shrinking diameters and possible explanations, In: *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2005.

[19] J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, Realistic, mathematically tractable graph generation and evolution, using Kronecker multiplication, In: *Proceedings of European Conference on Principles and Practice of Knowledge Discovery in Databases*, 2005.

[20] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, A. Tomkins, Geographic routing in social networks, *Proceedings of the National Academy of Sciences* **102** (2005) 11623-11628.

[21] T. Łuczak, P. Prałat, Protean graphs, *Internet Mathematics* **3** (2006), 21–40.

[22] S. Milgram, The small world problem, *Psychology Today* **2** (1967) 60-67.

[23] A. Mislove, M. Marcon, K. Gummadi, P. Druschel, B. Bhattacharjee, Measurement and analysis of online social networks, In: *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, 2007.

[24] M.E.J. Newman, J. Park, Why social networks are different from other types of networks, *Phys. Rev. E* **68**(3) 036122 (2003).

[25] Twitterholic. Accessed February 15, 2010. `http://twitterholic.com/`

[26] D.J. Watts, S.H. Strogatz, Collective dynamics of ‘small-world’ networks, *Nature* **393** (1998) 440-442.

[27] Wikipedia: List of social networking websites. Accessed February 15, 2010. `http://en.wikipedia.org/wiki/List of social networking websites`

[28] YouTube, Advertising and Targeting. Accessed February 15, 2010. `http://www.youtube.com/t/advertising targeting`