

Milgram-Routing in Social Networks

Silvio Lattanzi^{*}
Google, Inc.
76 Ninth Avenue
New York, NY
silviol@google.com

Alessandro Panconesi[†]
Sapienza University of Rome
113 Via Salaria
Rome, Italy
ale@di.uniroma1.it

D. Sivakumar[‡]
Yahoo!
4401 Great America Parkway
Santa Clara, CA 95054, USA
dsiva@yahoo-inc.com

ABSTRACT

We demonstrate how a recent model of social networks (“Affiliation Networks”, [21]) offers powerful cues in local routing within social networks, a theme made famous by sociologist Milgram’s “six degrees of separation” experiments. This model posits the existence of an “interest space” that underlies a social network; we prove that in networks produced by this model, not only do short paths exist among all pairs of nodes but natural local routing algorithms can discover them effectively. Specifically, we show that local routing can discover paths of length $O(\log^2 n)$ to targets chosen uniformly at random, and paths of length $O(1)$ to targets chosen with probability proportional to their degrees. Experiments on the co-authorship graph derived from DBLP data confirm our theoretical results, and shed light into the power of one step of lookahead in routing algorithms for social networks.

Categories and Subject Descriptors

G.3 [Mathematics of Computing]: Probability and statistics—*Stochastic processes*

General Terms

Theory

Keywords

Social Networks, Affiliation Networks, Milgram’s experiment

1. INTRODUCTION

Milgram’s six-degrees-of-separation experiment [28] and the fascinating small world hypothesis that follows from it have been rich sources of interesting research in recent years. In this landmark experiment, human subjects were asked to deliver a letter to a target person in a far-away city, following a simple rule: If they knew the target on a first name basis, they would deliver the letter; otherwise, they would pass

the letter along to a friend, with the same instructions. The surprising outcome was that a reasonably large fraction of the letters reached the target and moreover, they did so in very few hops. The success of Milgram’s experiments led to the fascinating small world hypothesis: take any two people in a social network, and they will be connected by a short chain of acquaintances. The extent to which the hypothesis is true is still actively debated. In this paper we give new experimental and theoretical results concerning Milgram’s experiment.

Empirical results: We perform an “in silico” replica of the experiment where a cognitive “space of interests” is navigated. In our experiment we consider a social network of co-authorships of computer science papers. Two people in this networks are “friends” if they are co-authors. We then extract a space of interests consisting of computer science topics. In simulating the experiment, we go from person to person by moving to the friend of the current person that has more interests in common with the target. To the best of our knowledge, this is the first time that the concept of an interest space is used in a pure digital replica. Previous studies such as [25, 30] made use of geographical proximity only (the move is toward the friend that is geographically closest to the target), an overly constrained rendition of the experiment.

Theoretical results: Motivated by the experiment above, we give a new evolutionary model that captures the small world phenomenon together with other important properties of real-world social networks. Among these there are dynamic properties, such as densification and shrinking diameter, and static properties like heavy-tailed distribution of popularity (number of friends). The main aspect of the model is its built-in “interest space,” a space of concepts that is navigable.

An important issue with Milgram’s small world hypothesis is the difficulty of its verification. Milgram’s painstaking work enabled him to collect data on a few hundreds of individuals. The advent of the internet has made it possible to perform large-scale replicas of the experiment, as in [6], or genuine “in silico” experiments, where there is no human participation—the experiment is made only with social data in digital form. Such “genuine” digital replicas are the main focus of this paper. One such instance is the study in [30]. A snapshot of the social networking site LiveJournal was downloaded to obtain a social network of roughly 15 million individuals. The experiment was simulated by picking source and target at random, and by moving toward the target according to geographical proximity (geo-greedy): from

^{*}Research done during an internship at Google when the author was in the Dipartimento di Informatica of Sapienza University

[†]The author is supported by a Google Research Award and a Yahoo! Faculty Award

[‡]Research conducted while author was at Google Inc.

Copyright is held by the International World Wide Web Conference Committee (IW3C2). Distribution of these papers is limited to classroom use, and personal use by others.

WWW 2011, March 28–April 1, 2011, Hyderabad, India.

ACM 978-1-4503-0632-4/11/03.

the current node X we move to the neighbor of X that is closest to the target. In another instance [25], the diameter of the social network of IM chat exchanges was estimated and found to be compatible with the small world hypothesis. These “cyber-replicas” have the obvious advantage that one can test the small-world hypothesis with millions of individuals. The main problem is how to make them realistic. The limitation of the approaches in [25, 30] is that they only take into account geographical or positional information, while it is clear that cognitive cues play a role in the original experiment. As mentioned, in this paper we present a cyber-replica of Milgram’s experiment in which a “space of concepts” is navigated. In simulating the experiment, we go from person to person by moving to the friend of the current person that has more interests in common with the target. Our experiments strongly reinforce two significant pieces of work in the sociology literature — the importance of *weak ties* [12] and the significance of the social status of the target node in Milgram’s experiment [17]. Finally, since our experiments are based on publicly available data, it should be possible for other researchers to replicate our work as well as derive additional insights underlying small-world routing.

There is a rich literature on stochastic models that reproduce several salient features of real-world social networks (e.g. power law distributions of popularity [5, 8], high clustering coefficient [26], densification and shrinking diameter [24]). There is, however, no model that seamlessly captures all of them. It would also be nice to have an evolutionary, as opposed to static, model of small world, where the network is evolving as new entrants join instead of being fixed. And of course, most interestingly, such a model should have a natural notion of “space of interests” that is navigable and that co-evolves with the social network. In this paper we do precisely this, by introducing a new evolutionary model that builds on the work in [21] and that captures nicely many salient sociological properties of real-world networks.

The model is based on *affiliation networks*, a concept first introduced in sociology by Breiger [4] and extended in [21]. An affiliation network is a bipartite graph with people on one side and interests on the other. The affiliation network comes with an associated *friendship graph* in which two people are friends if they share an interest. In the friendship graph people can also become friends because of “popularity” of one of the two parties, that is, preferential attachment is also in effect. In the original model people and interests join dynamically [21]. Each new node (person or interest) is a random perturbation of one pre-existing node. In our new model, a new interest that joins the network can be a perturbation of a mixture of pre-existing interests and, similarly, a new person joining the network will share a subset of the interests of several friends, as opposed to just one of them. Thus, this extension is more natural. While this is a small variation of the original model, the new model exhibits several interesting new properties (that are not known to be enjoyed by the original model). Our model is the first to exhibit simultaneously three different (sets of) properties of social networks: small world phenomena, evolutionary properties, and navigability of the interest space. In previous attempts, these features were somehow captured but separately. For instance, the models in [9, 10, 16, 35] deal with the small world phenomenon, but they are static and unable to explain evolutionary properties or even the

heavy-tailed distribution of popularity (number of friends). Furthermore they assume that every person knows the distance between its neighbors and the target, while we only assume that every person knows how similar interests are. There have been also some attempts to define and navigate an interest space instead of geographic information [15, 34] or to use a latent space of interests to define the friendship graph [31, 33]. But, again, these models are static (the number of nodes in the graph does not increase with time) and unable to explain evolutionary properties. In contrast, in our proposal all these different aspects come forth naturally from the same simple model.

Our enhanced model has several strong properties that are especially relevant for modeling small worlds, matching the experimental evidence from a quantitative point of view. The effective diameter of the friendship graph is bounded from above by a constant. This is compatible with the empirical observations of [25] where a very large social-network of hundreds of millions of nodes was analyzed, and its effective diameter found to be a very small number. When we analyze the actual working of Milgram routing in the friendship graph (not to be confused with the mere existence of short paths), we find that when source and target are chosen at random, their expected routing distance is $O(\log^2 n)$. The novelty here is that to find this short chain we navigate the interest space associated with the affiliation network, and not the friendship graph itself. When the target is chosen by popularity, i.e. with probability proportional to the numbers of friends, then the expected length of the chain can be upper bounded by a constant. This is in line with the experimental evidence with human subjects. It has been pointed out that the successful outcome of Milgram’s experiment could depend on the fact that the target was a person of high social status and had a profession that contributed even more than his status to establish and nurture many social connections [17]. Our model captures these features of the real world very nicely. Further, in accordance with the observation of Granovetter [12], the proofs of the upper bound for the diameter and the expected routing distance use heavily the presence of weak ties (i.e. preferential attachment edges in the model). To summarize, our analysis shows that our model incorporates not only basic structural facts of real-world networks, but can also explain some of their more nuanced features.

1.1 Related Work

We now overview the most relevant literature. Local routing algorithms have been intensely studied in the context of distributed systems. In this context there are some attempts to use the intuition behind the Milgram’s experiments to build new algorithms. Besides the work in [25, 30] that make use of social networks, other authors have replicated Milgram’s experiment in the real world [13] or by using email [2, 6]. These experimental findings are compatible with the small world hypothesis. The issue of attrition, the natural tendency of human subjects to drop out of the experiment, is analyzed in [11]. This social attrition introduces a bias in favor of short chains, because long chains tend to be interrupted before reaching the target. Taking this bias into account makes chains somewhat longer on average. Other interesting critiques to the Milgram’s experiment are presented in [17].

From a theoretical viewpoint, one of the first observations

that led to the interest in random graph models significantly different from the classical Erdős–Rényi models comes in the work of Faloutsos et al. [8], who noticed that the degree distribution of the Internet graph (the graph whose vertices are computers and whose edges are network links) is heavy-tailed, and roughly obeys a “power law,” that is, for some constant $\alpha > 0$, the fraction of nodes of degree d is proportional to $d^{-\alpha}$. Similar observations were made about the web graph (the graph whose vertices are web pages, and whose directed edges are hyperlinks among web pages) by Barabasi and Albert [3], who also presented models based on the notion of “preferential attachment,” wherein a network evolves by new nodes attaching themselves to existing nodes with probability proportional to the degrees of those nodes. Both works draw their inspiration and mathematical precedents from classical works of Zipf [36], Mandelbrot [27], and Simon [32]. Later Broder et al. [5] made a rich set of observations about the degree and connectivity structure of the web graph, and showed that besides power-law degree distribution, the web graph consisted of numerous dense bipartite subgraphs (often dubbed “communities”). Aiello et al. [1] and Kumar et al. [19] presented three models of random graphs that offer rigorous explanations for power-law degree distributions.

After the discovery of some surprising evolutionary properties such as densification and shrinking diameter in [24], several new models have been introduced [24, 22, 23] but none of them could really explain the new properties before the introduction of the *affiliation network* model [21]. The affiliation network model is the first model where interests have a crucial role and so is the first evolution model where it is possible to study Milgram’s experiment. In two previous papers [20, 29], the connectivity and the degree distribution of a static version of affiliation network model have been studied.

The works of Watts and Strogatz [35] and of Kleinberg [16] are the closest in spirit to ours in that they offer graph models that incorporate natural routing algorithms. In Kleinberg’s model, vertices reside in some metric space, and a vertex is usually connected to most other vertices in its metric neighborhood, and, in addition, to a few “long range” neighbors. He proved the remarkable result that the network has small diameter and easily discoverable paths iff the long-range neighbors are chosen in a specific way. Kleinberg’s models offer a nice starting point to analyze social networks, but because of its stylized nature, isn’t applicable in developing an understanding of the structure of real social networks. The other limitation of Kleinberg’s model is that it is static, and is not a model of graph evolution. For this reason several extensions of Kleinberg’s model have been introduced in order to study the problem starting from a different initial topology [10] or adding some constraint on the final degree distribution of the graphs [9].

2. OUR MODEL

The model that we consider in this paper is a variation of the one presented in [21]. In both models, two graphs evolve at the same time. The first is a bipartite graph, denoted as $B(P, I)$, that represents the affiliation network, with a set P of people on one side and a set of interests I on the other. An edge (p, i) represents the fact that p is interested in i . The second graph is a friendship network, denoted as $G(P, E)$, representing friendship relations within the same set P of

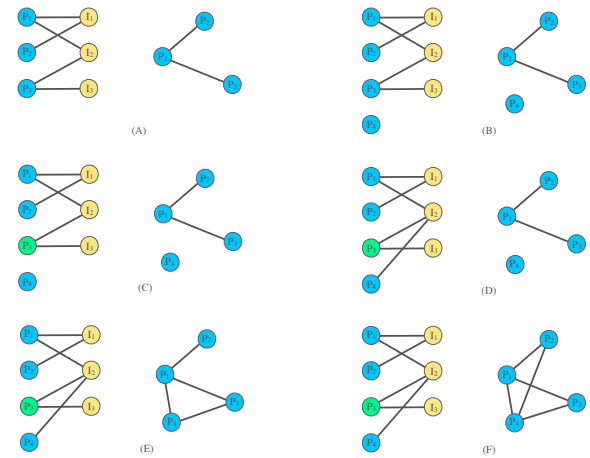


Figure 1: Insertion of a new person in the affiliation network and the social network derived from it. (A) The initial affiliation network and the related social graph. (B) Insertion of P_4 in the affiliation network. (C) P_4 selects as prototype P_3 . (D) P_4 copies a perturbation of the edges of P_3 . (E) The social graph is updated. (F) P_4 adds some preferential attachment edges in the social graph.

people. In this graph, people can be friends for two different reasons: if they share an interest or because of preferential attachment. Thus, G is the “folding” of B , plus a set of edges generated by preferential attachment. In [21] the graph B evolves as follows. When a new interest (resp. person) comes in, it selects a prototype node among the existing interests (resp. people) and copies it with a small perturbation. In this new version, when a new node joins B it can select more than one prototype. A new interest for example, will be a slightly perturbed mixture of a few existing interests, and a new person will be interested in a combination of interests of his/her friends. This new model seems more realistic and, from the technical point of view, it presents a few complications that make it a non straightforward extension of the previous one. More importantly, in this new version of the model it is possible to prove that it enjoys some interesting additional properties. Figure 1 describes the insertion of a new person in the affiliation network and the friendship network. Table 1 describes the model precisely. For readability, we present the two evolution processes separately even though the two graphs evolve together.

3. PRELIMINARIES

We say that an event occurs with high probability (whp) if it happens with probability $1 - o(1)$, where the $o(1)$ term goes to zero as n , the number of vertices, goes to ∞ . A random variable X is said to be heavy-tailed if $\lim_{x \rightarrow \infty} e^{\lambda x} \Pr[X > x] = \infty$ for all constants $\lambda > 0$.

DEFINITION 1. The graph $G(P, E)$ will be referred to as the friendship graph. An edge of G between two people that comes from the fact that they share an interest in B is called a folded edge.

| $B(P, I)$ | $G(P, E)$ |
|--|--|
| <p>Fix two integers k_1 and k_2, fix $k + s$ integers $\sum_{j=1}^{k_1} c_{p_j} = c_p$, $\sum_{j=1}^{k_2} c_{i_j} = c_i > 0$, and let $\beta \in (0, 1)$. At time 0, the bipartite graph $B_0(P, I)$ is a simple graph with at least $c_p c_i$ edges, where each node in P has at least c_p edges and each node in I has at least c_i edges.</p> <p>At time $t > 0$:</p> <p>(Evolution of P) With probability β:</p> <p><i>(Arrival)</i> A new node p is added to P.</p> <p><i>(Preferentially chosen Prototypes)</i> A set of nodes $p_1, \dots, p_{k_1} \in P$, with $k > 1$, are chosen as <i>prototypes</i> for the new node, with probability proportional to their degrees.</p> <p><i>(Edge copying)</i> c_{p_j} edges are “copied” from p_j, with $1 \leq j \leq k_1$ and $\sum_{j=1}^{k_1} c_{p_j} = c_p$; that is, c_{p_j} neighbors of p_j, denoted by $i_1, \dots, i_{c_{p_j}}$, are chosen uniformly at random (without replacement), and the edges $(p, i_1), \dots, (p, i_{c_{p_j}})$ are added to the graph.</p> <p>(Evolution of I) With probability $1 - \beta$, a new node i is added to I following a symmetrical process, adding c_i edges to i.</p> | <p>Fix three integers k_1, k_2 and s, fix $k + s$ integers $\sum_{j=1}^{k_1} c_{p_j} = c_p$, $\sum_{j=1}^{k_2} c_{i_j} = c_i > 0$, and let $\beta \in (0, 1)$. At time 0, $G_0(P, E)$ consists of the subset P of the vertices of $B_0(P, I)$, and two vertices have an edge between them for every neighbor in I that they have in common in $B_0(P, I)$.</p> <p>At time $t > 0$:</p> <p>(Evolution of P) With probability β:</p> <p><i>(Arrival)</i> A new node p is added to P.</p> <p><i>(Edges via Prototype)</i> An edge between p and another node in P is added for every neighbor that they have in common in $B(P, I)$ (note that this is done after the edges for p are determined in B).</p> <p>(Edges via evolution of I)</p> <p>With probability $1 - \beta$:</p> <p>A new edge is added between two nodes p_1 and p_2 if the new node added to $i \in I$ is a neighbor of both p_1 and p_2 in $B(P, I)$.</p> <p>(Preferentially Chosen Edges) A set of s nodes p_{i_1}, \dots, p_{i_s} is chosen, each node independently of the others (with replacement), by choosing vertices with probability proportional to their degrees, and the edges $(p, p_{i_1}), \dots, (p, p_{i_s})$ are added to $G(P, E)$.</p> |

Table 1: Description of the evolving model.

4. BASIC PROPERTIES OF THE MODEL

In this section we show that the properties of the original model in [21] are also enjoyed by the new model. We begin by defining the concept of *effective diameter* that intuitively measures the largest distance between “almost all” pair of nodes in a graph.

DEFINITION 2. [Effective Diameter] For $0 < q < 1$, the q -effective diameter is the minimum d_e such that, for at least a q fraction of the node pairs, the length of the shortest path between the pair is at most d_e .

Then we define the *core* and *hubs*. Intuitively, they define the popular interests and the people that are connected to them.

DEFINITION 3. [Core and hubs] Let $d(v)$ be the degree of v . A set of interests $C \subseteq I$ is an $\alpha\epsilon$ -core of an affiliation network $B(P, I)$ if $d(v) \geq \alpha n^\epsilon$ for all $v \in C$. The hubs are the people in P at distance one from C .

In what follows we will refer to $\alpha\epsilon$ -cores simply as cores. We now list a set of properties that our model shares with the original affiliation network of [21]. The proofs are similar and omitted from this extended abstract. The properties are the heavy tailed distribution of degrees in both $B(P, I)$ and $G(P, E)$, and densification and shrinking diameter of $G(P, E)$.

THEOREM 1. [General properties of the model]

(1) Given an affiliation network $B(P, I)$, the degree sequence of nodes in P (resp. I), almost surely when $n \rightarrow \infty$, follows a power law distribution with exponent

$$\alpha = -2 - \frac{c_p \beta}{c_i(1 - \beta)}$$

(resp. $\alpha = -2 - \frac{c_i(1 - \beta)}{c_p \beta}$), for every degree smaller than n^γ

where

$$\gamma < \frac{1}{4 + \frac{c_p \beta}{c_i(1 - \beta)}} \left(\gamma < \frac{1}{4 + \frac{c_i(1 - \beta)}{c_p \beta}} \right)$$

(2) The degree distributions of the graphs $G(P, E)$ is heavy-tailed with high probability.

(3) The number of edges in $G(P, E)$ is $\omega(n)$ with high probability.

(4) The q -effective diameter of $G(P, E)$ shrinks or stabilizes after time ϕn with high probability, for any constant $0 < \phi < 1$ and for any constant $0 < q < 1$.

Now we state two technical lemmas that we will use in the following sections. The proofs are omitted from this extended abstract.

LEMMA 2. Let $1 > \epsilon > 0$ be any constant and let v be a node in $B(P, I)$ with degree $g(n)$ at time ϵn , with $g(n) \in \Omega(\log^2 n)$. Then, with high probability, v 's degree at time n is smaller than $C \cdot g(n)$, for some constant $C > 0$. Furthermore, if a node v has degree $o(\log^2 n)$ at time ϵn or it is inserted after time ϵn , then the final degree of v is in $o(\log^2 n)$ with high probability.

LEMMA 3. With high probability, any node of P inserted after time ϕn , for any constant $\phi > 0$, will be connected to a hub via a preferential attachment edge. Further we have that, with high probability for $t > \phi n$:

$$V(\text{hubs}, t) = \sum_{v \in \text{hubs}} d_G(v) \in \Omega \left(t^{1+\gamma \left(1 - \frac{c_i(1-\beta)}{c_p \beta} \right)} \right)$$

and

$$V(G/\text{hubs}, t) = \sum_{v \notin \text{hubs}} d_G(v) \in \Theta \left(t^{1+\epsilon \left(1 - \frac{c_i(1-\beta)}{c_p \beta} \right)} \right)$$

Where γ and ϵ are two constants such that $\gamma > \epsilon$ and $V(S, t)$ is the volume of the node in S at time t .

5. THE CRUCIAL ROLE OF WEAK TIES

In this section we study the effective diameter of $G(P, E)$ and show that it is bounded by a constant (it is unknown if this property holds in the original affiliation network model). This property is a consequence of the co-existence of folded and preferential-attachment edges. Several studies have shown that links in a social network are of two types, local and long-range, also called weak, ties [12]. Weak ties have several important structural properties, for instance they form bridges between different communities and, in particular, they are the crucial ingredient that makes small worlds possible.

In our model folded edges are local, for they connect people within a community of shared interests, while preferential attachment edges are the weak (or long-range) ties [12, 16]. Note that, in accordance with the previous literature and sociological intuition, in our model weak ties are very few compared to folded edges. In this section we show that weak ties play another interesting structural function that is in accordance with the empirical evidence: weak ties are crucial to bound the effective diameter of the friendship graph by a constant.

Our proof also uses in a fundamental way the presence of hubs. This might seem in contrast with the results in [6] where the authors suggest that their role is not relevant. A possible explanation is that they consider only the degree induced by the explored paths, and thus consider only a subgraph of the social network. Thus it is possible that in their experiments a high degree node seems to have small degree just because only few messages passed through him. In our proof we consider the real degree of a node. We also note that our results are in line with the original findings of Milgram [27] and with our experiments.

THEOREM 4. *For every $q < 1$, there is a constant Δ_q such that the q -effective diameter of $G(P, E)$ is bounded from above by Δ_q .*

The next lemma (proof omitted) on the distance between nodes in the core of the affiliation network is crucial in the proof of Theorem 4.

LEMMA 5. *Let C be the core of the affiliation network $B(P, I)$. There exists a constant d , independent of n , such that, with high probability, the distance in the affiliation network between any pair of nodes in the core is at most d .*

COROLLARY 6. *Any two hubs are at constant distance in $G(P, E)$ and $B(P, I)$, with high probability.*

PROOF. (of Theorem 4) Recall that from Lemma 3 we have that all nodes in P inserted after time ϕn , for any $\phi > 0$, will have at least one preferential attachment edge incident to a hub, with probability $1 - o(1)$. Now, let X_i be a random variable such that:

$$X_i = \begin{cases} 1 & \text{if } i \text{ has a hub in its neighborhood} \\ 0 & \text{otherwise} \end{cases}$$

The number of nodes that have at least one hub in their neighborhood is $\sum_{i=1}^n X_i \geq \sum_{i=\phi n}^n X_i$. From Lemma 3 it follows that $E \left[\sum_{i=\phi n}^n X_i \right] \geq (1 - c)n$, for any constant $c >$

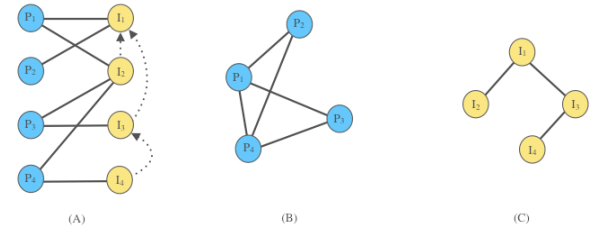


Figure 2: An affiliation network(A) and the induced social network(B) and hierarchy of interests(C). The dotted lines from a to b in (A) represent that b is the prototype of a .

ϕ . Observe that each X_i satisfies the Lipschitz condition with d_i equal 1. So by standard concentration results [7] we have that $\sum_{i=\phi n}^n X_i \geq (1 - c')n$, for any constant $c' > c$. Hence the claim follows from Corollary 6. \square

6. LOCAL ROUTING IN INTEREST SPACE

In this section we analyze the performance of Milgram routing in our model. It is clear that in Milgram's experiment cues other than geographic distance play a role. For instance, the target was defined not only by a location but, crucially, by a profession. Therefore, if one wants more realistic models a more nuanced version of proximity must be used. In this section we show that our model has a natural "space of interests" that is associated with the affiliation network that is navigable. We note that it is not known if the original affiliation network model enjoys the same property [21].

Two more aspects make the following analysis interesting in our opinion. This is the first study of the performance of local routing algorithm with an evolving model. Furthermore, ours is the first model that can explain Milgram's experiment if we assume some constant attrition, as suggested in [11] (i.e. in this case only paths of constant length can be observed with high probability).

We start by defining a notion of distance between interests. In order to do so we define the *prototype graph* $G(I, \tilde{E})$. The nodes of the prototype graph are the interests in the affiliation network, and two interest i_1, i_2 have an edge between them if i_1 has been selected as a prototype for i_2 or vice versa. Furthermore, we have that two initial interests i' and i'' contained in the graph $B_0(P, I)$ are connected if there is a person in $B_0(P, I)$ that is interested in (connected to) both. Thus, the prototype graph consists of a clique of the initial interests and of links connecting nodes to their prototypes. In Figure 2 it is shown an example of affiliation network with the induced friendship network and the prototype graph.

DEFINITION 4. [Distance between interests] *For two nodes $i_1, i_2 \in I$, we define the distance between i_1 and i_2 as the shortest (hop) distance between the two nodes in the prototype graph. Further, we define the interest distance between two people p_1 and p_2 as the smallest distance between any interest of p_1 and any interest of p_2 .*

In our analysis we assume that every person is able to assess the distance between any two interests. In practice we are

assuming that every person is able to compute the similarity between any two interests, in order to decide which friend is closest to the target. This natural assumption is made, perhaps implicitly, in every previous navigation model. For example in [16] a node is always able to select the neighbor closest to the target in the metric space.

We define our routing algorithm as follows.

DEFINITION 5. [Local Routing algorithm]

In each step the message holder u performs the following:

- (1) *If the destination is a neighbor of u , the message is forwarded to it.*
- (2) *Otherwise, u forwards the message to the neighbor that minimizes the interest distance to the destination.*

We start by proving a basic property of our algorithm.

LEMMA 7. *In every step of the local routing algorithm, either the interest distance between the message holder and the destination is reduced or the message is delivered to the target.*

PROOF. If the message holder knows the target the lemma is true by definition 5. Otherwise let v be any interest of the message holder and let $w(v)$ be an interest connected to v in the *prototype graph* but with smaller distance from the target. Note that $w(v)$ always exists because the graph is connected.

There are two cases: (i) if $v, w(v) \in B_0(P, I)$ then there is a person in $B_0(P, I)$ interested to both $v, w(v)$; (ii) if v is a prototype of $w(v)$ (or, symmetrically, vice-versa) then v and $w(v)$ have a neighbor in common in $B(P, I)$ by definition of the evolving process. In any case, for any interest v of the message holder in the people graph, there is a person interested in both v and $w(v)$. It follows that in the neighborhood of the message holder, for any interest v , there is a person interested in $w(v)$. So using the local routing algorithm it is always possible to forward the message to the neighbor closest to the target, and the claim follows. \square

We now show that for most source-destination pairs it is possible to route the message within a constant number of steps, provided that the destination is selected with a probability that is proportional to its degree, i.e. its “popularity” in the social network. This result is in accordance with the analysis of Milgram’s experiment done by Kleinfeld [17], who pointed out that a successful outcome crucially depends on the social status of the target¹.

THEOREM 8. *Let $c_i < \frac{\beta}{1-\beta} c_p$. If the destination is selected with probability proportional to its degree and the source is selected uniformly at random then, with probability $(1 - \phi - o(1))$, for any constant $\phi > 0$, the local routing algorithm delivers the message in constant many steps.*

PROOF. Let v be the destination. We first prove that with probability $1 - o(1)$ v is a hub. Recall that the *volume*

¹This point is in contrast with the claim in [6] but we recall Kleinfeld’s observation in [17]. She wrote “Take the selection of the sample. I found in the archives the original advertisement recruiting subjects for the Wichita, Kansas study. This advertisement was worded so as to attract not representative people but particularly sociable people proud of their social skills and confident of their powers to reach someone across class barriers.” Apart from this skepticism there are experiments suggesting that social barriers can actually hinder Milgram’s local routing [18].

of a vertex is its degree, and that the volume of a set of vertices is the sum of their volumes. Let $hubs$ denote the set of hubs. Let $V(hubs, t)$ be the total volume of the hubs at time t , and $V(G/hubs, t)$ the total volume of the rest of the graph at time t . As shown in Lemma 3 we have that, for $t > \phi n$:

$$V(hubs, t) = \sum_{v \in hubs} d_G(v) \in \Omega \left(t^{1+\gamma \left(1 - \frac{c_i(1-\beta)}{c_p\beta} \right)} \right)$$

and

$$V(G/hubs, t) = \sum_{v \notin hubs} d_G(v) \in \Theta \left(t^{1+\epsilon \left(1 - \frac{c_i(1-\beta)}{c_p\beta} \right)} \right)$$

Where $\gamma > \epsilon$. Thus, when the destination is selected with probability proportional to its degree, with probability $1 - o(1)$, it will be a hub. In addition, Lemma 5 implies that two hubs are within constant distance also in the interest space. So, by Lemma 7, it holds with high probability that if a message reaches a hub it will need only a constant additional number of steps to reach every other hub using the local routing algorithm.

Now note that Lemma 2 implies that all the hubs are inserted before time ϕn with high probability, for every constant $\phi > 0$. Further, by Lemma 3 every node inserted after time ϕn will be connected to a hub with probability $1 - o(1)$. To summarize, with probability $(1 - \phi - o(1))$, the destination is a hub and the source has a hub in its neighborhood. It follows that the local routing algorithm will deliver a message in a constant number of rounds, with probability at least $(1 - \phi - o(1))$. \square

We now consider a different setting. Suppose that we expand the interests of the destination in such a way that they include the interest of its neighbors. We call this case the *expanded interests setting*. This is an attempt to capture the additional knowledge that human subjects have about the destination, apart from its personal information. This is interesting because it captures some features of the original experiment. For instance, in the first experiment presented by Milgram in [28], the sources knew also that the target was married to a divinity student in Cambridge, MA.

In this setting we can prove the following. The proof is similar to the proof of the previous Theorem and omitted for lack of space.

THEOREM 9. *Let $c_i < \frac{\beta}{1-\beta} c_p$. In the expanded interests setting when source and destination are selected uniformly at random then, with probability $(1 - 2\phi - o(1))$, the local routing algorithm will route the message in constant many steps, for every constant $\phi > 0$.*

Now we study the most general case, when source and target are chosen adversarially and we do not extend the interest space of the destination, in this setting we are able to show the following upper bound on the running time of the local routing algorithm.

THEOREM 10. *If $c_i < \frac{\beta}{1-\beta} c_p$ then, for any source and any destination, the local routing algorithm routes the message within $O(\log^2 n)$ steps with high probability.*

PROOF. To prove the result we will bound the diameter of the interest prototype graph. By Lemma 7 the diameter is an easy upper bound for the delivery time of local routing.

We will show that, with high probability (whp), the diameter of the *prototype graph* is $O(\log^2 n)$.

The general idea of the proof is to divide the random process in $O(\log n)$ macro-phases, and to show that in each macro-phase the probability that diameter increases by $\omega(\log n)$ is $o\left(\frac{1}{\log n}\right)$. Thus, the diameter is $O(\log^2 n)$ whp.

Let us divide the evolving process in $O(\log n)$ phases. In phase zero we group the first $600 \log n$ steps. Phase one is from the end of phase zero to step $\lfloor (1 + \epsilon) \log n \rfloor$, for a small constant $\epsilon > 0$. Phase two is up to step $\lfloor (1 + \epsilon)^2 \log n \rfloor$. In general, phase i starts after the end of phase $i - 1$ and ends at step $\lfloor (1 + \epsilon)^i \log n \rfloor$.

Let us now consider a generic phase $t > 0$. Let $T = (1 + \epsilon)^t 600 \log n$. First, we want a bound on the number of edges in the affiliation network $B(P, I)$ at the beginning of each phase. Let A_t be the random variable that counts the number of edges at the beginning of phase t . We have that $E[A_t] = (\beta c_p + (1 - \beta) c_i) T$. By the Chernoff bound,

$$\Pr \left[|E[A_t] - A_t| > \frac{1}{10} E[A_t] \right] \leq \exp \left(-\frac{E[A_t]}{300} \right) \leq \frac{1}{n^2}.$$

Using the union bound for the number of macro-phases, it follows that at the beginning of each phase t , $\frac{9}{10} E[A_t] \leq A_t \leq \frac{11}{10} E[A_t]$ with high probability. In the rest of the proof we will assume that $\frac{9}{10} E[A_t] \leq A_t \leq \frac{11}{10} E[A_t]$.

To get a bound on the diameter, we start by studying the two following events:

- $\xi_1(j) = \{\text{interest } j, \text{ inserted in phase } t, \text{ of degree } c_i \text{ selected in a step during phase } t \text{ as a prototype for the first time}\}$
- $\xi_2(j) = \{\text{interest } j, \text{ inserted in phase } t, \text{ of degree } c_i \text{ increases its degree in a step during phase } t\}$

First notice that from the definition of the evolving process, we have that $\Pr[\xi_1(j)] \leq \frac{c_i}{A_t} \leq \frac{10c_i}{9T}$.

To bound $\Pr[\xi_2(j)]$, recall that interest j has degree c_i , so there are c_i people interested in it. Denote them as p_1, p_2, \dots, p_{c_i} . Now, if j increases its degree, it must be because a new person joins the graph and copies the interest j from one of the person interested to it, p_1, p_2, \dots, p_{c_i} . This happens with probability:

$$\Pr[\xi_2(j)] \leq \sum_{i=1}^{c_i} \frac{10d_i}{9T} \left(1 - \left(1 - \frac{1}{d_i} \right)^{c_p} \right)$$

Using calculus, it is possible to see that this probability is maximized when $d_1 = \dots = d_{c_i} = T$. Thus

$$\Pr[\xi_2(j)] \leq c_i \left(1 - \left(1 - \frac{1}{T} \right)^{c_p} \right) \leq c_i \left(1 - e^{-\frac{c_p}{T}} \right) \leq \frac{c_p c_i}{T}$$

So $\Pr[\xi_1(j) \vee \xi_2(j)] \leq 2 \frac{c_p c_i}{T}$. Let us define $\xi(j) := \xi_1(j) \vee \xi_2(j)$.

Now we can compute the probability that in phase t the diameter of the *prototype graph* increases by more than C , with $C > e$. Let us call this event τ_C . Note that if τ_C holds, a sequence of C new interests is added in phase t , increasing the diameter of the prototype graph of C . In order for this event to hold ξ has to occur at least C times in a phase. So

we can upper bound τ_C as follows.

$$\begin{aligned} \Pr[\tau_C] &\leq (\# \text{ of steps in a phase}) \cdot (\# \text{ of new nodes in a phase}) \cdot P[x_i(j) \text{ holds for node } j \text{ of degree } c_i] \\ &\leq \lceil \epsilon T \rceil \left(\frac{\lceil \epsilon T \rceil}{C} \right) P[x_i \text{ holds for a node } j \text{ in a step}] \\ &\leq \lceil \epsilon T \rceil \left(\frac{\lceil \epsilon T \rceil}{C} \right) \left(2 \frac{c_p c_q}{T} \right)^C \\ &\leq \lceil \epsilon T \rceil \left(1 + \frac{C}{\epsilon T - C} \right)^{\epsilon T - C} \left(\frac{\epsilon}{C} \right)^C (2c_p c_q)^C \\ &\leq \lceil \epsilon T \rceil e^C \left(\frac{\epsilon}{C} \right)^C (2c_p c_q)^C \\ &< \lceil \epsilon T \rceil (2\epsilon c_p c_q)^C, \end{aligned}$$

where in the third inequality we use Stirling's approximation. Therefore the probability of τ_C decreases geometrically with C .

Finally, let us compute the probability that the final diameter is greater than $K = k \log^2 n$. After the first phase the diameter is at most $600 \log n$, so we can bound the previous probability as the probability that the diameter increases by at least $(k - 600) \log n$ after phase 1. Hence

$$\begin{aligned} \Pr[D \geq k \log n] &\leq \sum_{\substack{k_2, k_3, \dots, k_{\log(1+\epsilon)n} \\ \sum_{i=2}^{\log(1+\epsilon)n} k_i = K - 600 \log n}} \Pi_{i=2}^{\log(1+\epsilon)n} \Pr[\xi_{k_i}] \\ &\leq \left(\log_{(1+\epsilon)} n \cdot (K - 600 \log n) \right) \cdot \left(T^{\log n} \cdot (2\epsilon c_p c_q)^{K - 600 \log n} \right) \\ &\leq \left(\log_{(1+\epsilon)} n \cdot (K - 600 \log n) \right) \cdot \Theta \left(n^{\log n} \cdot n^{-k \log n} \right) \\ &\in o(1) \end{aligned}$$

Thus by choosing a large enough k the claim follows. \square

7. EXPERIMENTS

Our mathematical model of social networks, building on the affiliation network model, suggests natural decentralized routing algorithms in social networks. Namely, given a source vertex s and a target vertex t , identify the interests of s and t in the underlying affiliation network and identify the neighbor of s whose interests are closer to that of t (with respect to the hierarchy of interests implied by the prototype selection step). Inspired by this, one can define natural algorithms that perform decentralized routing in real-world social networks by suitably approximating the process of navigating the interest hierarchy. In this section, we do precisely this, and report our findings based on simple experiments with a modestly-sized social network.

Our social network consists of *authors* as nodes and edges defined by *co-authorship* of one or more articles. We downloaded a copy of the DBLP database of computer science papers, a DB of roughly 735,000 authors and 1.24M articles, and constructed the co-authorship graph with about 4.63M edges (for an average degree of roughly 6.7 co-authors per node). On this network, we randomly selected about 575 pairs of source–target pairs and attempted to construct paths between them. The largest connected component in this network has roughly 80% of the vertices, with the rest of the vertices in very small isolated components, so that the probability that two randomly selected nodes belong to

the largest connected component is roughly 64%. The mean length of the shortest path between nodes in this component is roughly 6.3 (with a median length of 6).

Notice that in this way, we construct an affiliation network where two authors are friends if they coauthor a paper, now we have to infer a metric on the interest in order to route the messages. Unfortunately this is not easy, because there is not a clear definition of closeness between papers and all the standard classification system for the papers are too poor for our purpose. To overcome this difficulty we define the interest space not as the set of papers but as the set of bigrams and unigrams contained in the title of the paper.

In particular we begin by segmenting article titles into one-word and two-word sequences (unigrams and bigrams) after suitably eliminating stopwords that occur commonly ('and', 'the', etc.). For instance, the title "*Small world experiments for everyone*" generates four unigrams — 'small', 'world', 'experiments', and 'everyone', and two bigrams — 'small world', 'world experiments'. Both bigrams and unigrams are treated as interests, with the latter of a more generic kind; for instance, the unigram 'physics' is somewhat general, whereas the bigram 'particle physics' is much more specific. In this fashion, for every author, their *interest profile* is identified; specifically, for author a and interest i , we define $s(i, a)$ to be the *strength* of interest i for author a , and is defined as the number of occurrences of interest (unigram/bigram) i within author a 's publications.

To simulate Milgram's experiment, our basic algorithm operates as follows: if we are currently at node x , we move to the neighbor y of x whose interest profile is closest to the target t , where the measure of proximity of y to t is computed according to the formula

$$\text{proximity}(y, t) = \sum_{\text{Interest } i} \frac{s(i, y)s(i, t)}{p(i)},$$

where $p(i)$ denotes the overall popularity of interest i , defined by $p(i) = \sum_a s(i, a)$. If there is no neighbor with non-zero proximity, we either declare failure, or in a variation of the experiment, proceed greedily to the neighbor of highest degree.

The most basic variant of the algorithm outlined insists that the proximity measure strictly increase in each step of the routing: this version is called LOCAL-MONOTONE, and the version without this restriction is called LOCAL. The next variation we consider is to allow one step of 'lookahead', where we not only evaluate neighbors of x , but also evaluate neighbors of neighbors of x , and route through the neighbor whose neighbor achieves the highest proximity to the target; this idea of 'lookahead', very common in computer science, captures the belief that in real social networks, one not only has knowledge about their friends, one often has partial knowledge about friends-of-friends. The corresponding non-monotone and monotone variations are called, respectively, LOOKAHEAD and LOOKAHEAD-MONOTONE. In a third variation, we allow the algorithm the knowledge not only of the target's interests, but also those of its neighbors; this is a 'reverse' and limited form of lookahead, and has precedent in Milgram's experiment, where the sources had the knowledge that the target was the wife of a student of divinity in Cambridge, Mass. This is naturally aimed at routing to hard-to-reach destinations by augmenting the algorithm with extra information. The corresponding vari-

ations of the four algorithms described above are LOCAL-EXPAND, LOCAL-MONOTONE-EXPAND, and so on.

Figure 3, 4 report the percentage of successful chains for the eight variations of the decentralized routing algorithm we studied. For reference, we compare the performance of the decentralized routing algorithms to that of the omniscient algorithm that has full information about the network structure and employs a standard 'shortest path' computation. The 'success percentage' in Figure 3, 4 is the percentage of source-target pairs successfully routed, divided by 0.64 (which is the fraction for this omniscient algorithm). The results are presented in four groups, each corresponding to one value of a parameter called τ , which restricts the sampling of the target nodes to be uniform among all nodes of degree at least τ ; this is done to explore the role of the centrality of the target in determining the success of decentralized routing.

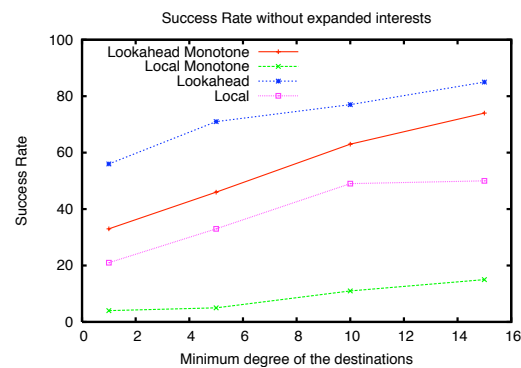


Figure 3: Success Rate without extended interests.

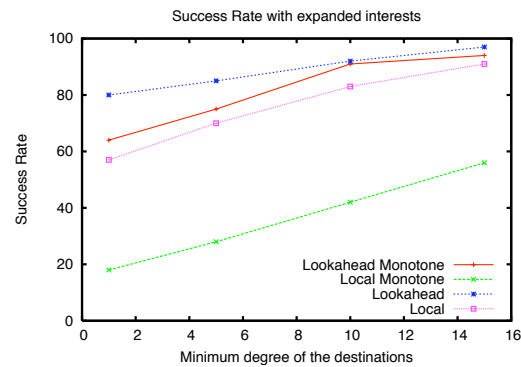


Figure 4: Success Rate with extended interests.

We briefly highlight some salient observations based on Figures 3, 4, 5 and 6 and other related experiments.

(1) Navigation based on interests is an extremely powerful paradigm; the success of the basic algorithm LOCAL in achieving 21% successful routing is, a priori, unexpected, given how crude our construction of the interest space is. In particular the previous replicas of the small-world experiment had always lower successful rate [6, 30].

(2) Adding even one of two natural cues to local routing (either expanding the interests of the target or adding a step of lookahead) is enormously powerful — with each cue

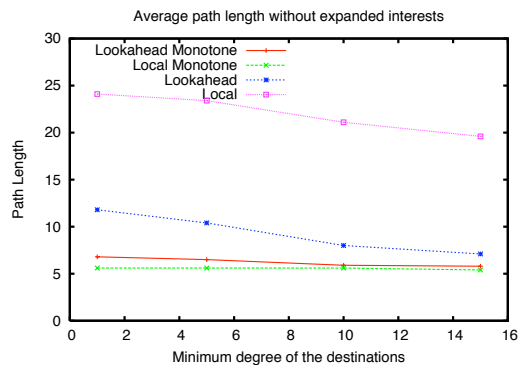


Figure 5: Average path length without extended interests.

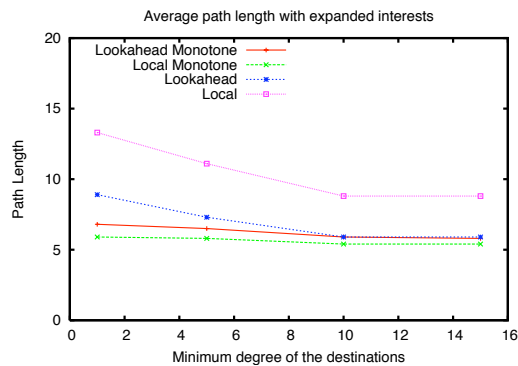


Figure 6: Average path length with extended interests.

raising the success rate to about 57%, and reducing the path length from about 24 to about 12.

(3) Adding both interest expansion and lookahead results in 80% successful routing, with extremely short paths (a median path length of 7).

(4) Insisting on monotonically better proximity to the target's interests typically reduces success rate, but significantly improves the length of the path constructed, for each of the four variations of the algorithm.

(5) Picking the target from a distribution that is restricted to targets of certain minimum degree dramatically improves the success rate and path length for decentralized routing algorithms. While this restriction might appear strange, this captures the idea the even modestly 'well-connected' nodes are significantly easier to reach than completely isolated ones. When we place a minimum degree restriction of 15 (recall that the average degree is only 6.7), the best algorithm achieves 97% success rate and produces paths almost as short as the shortest possible! Even the simplest of algorithms, LOCAL, succeeds on 50% of the cases — this reinforces the argument made by Kleinfeld, who, analyzing Milgram's experiments, suggests that the success of the routing depends, to some extent, on the fact that the target was not an isolated person but one well-connected in terms of geographic location, employment, social status, etc.

(6) Besides the results plotted in Figures 3–6, we also explored the importance of *core* nodes, and more generally, the role of *weak ties* in social routing. Specifically, we iden-

tified the node of highest degree along successful paths, and computed the average and median of its degree. While the average degree of author nodes in the co-authorship network is 6.7, the average and median values of the degree of the node with the most connections along shortest paths are, respectively, 133 and 163. For LOOKAHEAD-EXPAND, our most successful decentralized routing algorithm, these values are, respectively, 189 and 228. These findings reinforce the arguments of Granovetter [12] concerning the strength of weak ties, as well as our analytical results proving the importance of core nodes for decentralized routing.

8. REFERENCES

- [1] W. Aiello, F. Chung and L. Lu, "Random Evolution in Massive Graphs". In FOCS'01, 42 (2001), 510-520.
- [2] L. Adamic and E. Adar, "How to search a social network". Social Networks, 27 (3) (2005), 187-203.
- [3] R. Albert and A.-L. Barabasi. "Emergence of scaling in random networks". Science, 286 (1999), 509-512.
- [4] R. L. Breiger, "The Duality of Persons and Groups". Social Forces, University of North Carolina Press, 1974.
- [5] A. Z. Broder, S. R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins and J. Wiener, "Graph structure in the web". In WWW'00, 9 (2000), 309-320.
- [6] P.S.Dodds, R.Muhamad and D.J.Watts, "An experimental study of search in global social networks". Science, 301(5634) (2003), 827-829.
- [7] D. Dubhashi and A. Panconesi, "Concentration of Measure for the Analysis of Randomised Algorithms". Cambridge University Press, 2009.
- [8] M. Faloutsos, P. Faloutsos and C. Faloutsos, "On power-law relationships of the Internet topology". In the conference on Applications, technologies, architectures, and protocols for computer communication, (1999), 251-262.
- [9] P. Fraigniaud and G. Giakkoupis, "The effect of power-law degrees on the navigability of small worlds". In PODC'09, 28 (2009), 240-249.
- [10] P. Fraigniaud, and G. Giakkoupis, "On the searchability of small-world networks with arbitrary underlying structure". In STOC10, 42 (2010), 389-398.
- [11] S. Goel, R. Muhamad and D. J. Watts, "Social search in "Small-World" experiments". In WWW09, (2009), 701-710.
- [12] M. Granovetter, "The Strength of Weak Ties". American Journal of Sociology, 78(6) 1973, 1360-1380.
- [13] P. Killworth and H. Bernard, "Reverse small world experiment". Social Networks, 1 (1978), 159-192.
- [14] V. Klee and D. Larman, "Diameters of random graphs". Canad. J. Math., 33 (1981), 618-640.
- [15] J. Kleinberg, "Small-World Phenomena and the Dynamics of Information". Advances in Neural Information Processing Systems (NIPS'01), 14 (2001), 431-438.
- [16] J. Kleinberg, "The small-world phenomenon: An algorithmic perspective". In STOC'00, 32 (2000), 163-170.
- [17] J. Kleinfeld, "Could it be a big world after all?". Society, 39 (2002), 61-66.

- [18] C. Korte, and S. Milgram, “Acquaintance links between White and Negro populations: Application of the small world method”. *Journal of Personality and Social Psychology* 15(2), 101-108.
- [19] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins and E. Upfal, “Stochastic models for the web graph”. In *FOCS’00*, 41 (2000), 57-65.
- [20] M. Karoński, E. R. Scheinerman and K. B. Singer-Cohen, “On Random Intersection Graphs: The Subgraph Problem”, *Combinatorics, Probability and Computing*, 8(1–2), 2006, 131–159.
- [21] S. Lattanzi and D. Sivakumar, “Affiliation Networks”. In *STOC’09*, 41 (2009), 427-434.
- [22] J. Leskovec, L. Backstrom, R. Kumar and A. Tomkins, “Microscopic evolution of social networks”. In *KDD’08*, 14 (2008), 462-470.
- [23] J. Leskovec, D. Chakrabarti, J.M. Kleinberg and C. Faloutsos, “Realistic, Mathematically Tractable Graph Generation and Evolution, Using Kronecker Multiplication”. In *PKDD’05*, (2005), 133-145.
- [24] J. Leskovec, J. Kleinberg and C. Faloutsos, “Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations”. In *KDD’05*, 11 (2005), 177 - 187.
- [25] J. Leskovec and E. Horvitz, “Planetary-scale views on a large instant-messaging network”. In *WWW’08*, 17 (2008), 915-924.
- [26] J. Leskovec, K. Lang, A. Dasgupta and M. Mahoney. “Statistical Properties of Community Structure in Large Social and Information Networks”. In *WWW’08*, 17 (2008), 695-704.
- [27] B. Mandelbrot, “An informational theory of the statistical structure of languages”, *Communication Theory*, (1953), 486-502.
- [28] S. Milgram, “The Small World Problem”. *Psychology Today*, 2 (1967), 60-67.
- [29] M. E. Newman, “Properties of highly clustered networks”, *Phys Rev E Stat Nonlin Soft Matter Phys*, 68(2), 2003.
- [30] D. L. Nowell, J. Novak, R. Kumar, P. Raghavan and A. Tomkins, “Geographic Routing in Social Networks”. *National Academy of Sciences*, 33(102) (2005), 11623-11628.
- [31] A. E. Raftery, M. S. Handcock and P. D. Hoff, “Latent space approaches to social network analysis”, *J. Amer. Stat. Assoc.*, 15(460), 2002
- [32] H. Simon, “On a class of skew distribution functions”. *Biometrika*, 42 (1955), 425-440.
- [33] P. Sarkar and A. W. Moore, “Dynamic social network analysis using latent space models”, *ACM SIGKDD Explorations Newsletter*, 7(2), 2005, 31-40.
- [34] D. J. Watts, P. S. Dodds and M. E. J. Newman, “Identity and Search in Social Networks”. *Science*, 296 (2002), 1302-1305.
- [35] D. Watts and S. Strogatz, “Collective dynamics of small-world networks”. *Nature*, 393(6684) 1998, 409-410.
- [36] G. K. Zipf, “Human Behavior and the Principle of Least Effort”. Addison-Wesley, 1949.