MIPRO 2014, 26-30 May 2014, Opatija, Croatia

Regular Equivalence in Informed Network Search

Denis Helic
Knowledge Management Institute
Graz University of Technology
Graz, Austria
Email: dhelic@tugraz.at

***Abstract*—Search in networks is defined as a process in which an agent hops from one network node to another by traversing network links in search for given nodes. The simplest example of network search is a random walk where the agent selects a link uniformly at random from all outgoing links of the current node. On contrary, in an informed search the agent possesses (partial) background knowledge of the network. This background knowledge steers the agent's decisions when selecting the next link to traverse. The background knowledge of the network can be represented as a similarity matrix with similarities between pairs of nodes known to an agent. This matrix can be calculated in various ways in order to model various search scenarios or to best fit needs of an application. For example, similarities based on node degrees or some external information about the nodes have been commonly used in the past. In this paper we evaluate the measures that capture regular equivalence of nodes in a network with respect to their suitability as a similarity metric to inform search in networks. In particular we are interested in the properties of Katz similarity for this task.**

I. Introduction

Search in a network is a dynamical process in which an agent moves along the links between the nodes in search for a given node or a given set of nodes. Depending on the link selection mechanism we distinguish between a random walk and an informed search. In a random walk an agent selects the next link uniformly at random from the available links (the links going out from the agent's current node). In an informed search the agent possesses some background knowledge of the network, which guides it towards selecting specific links. In other words, the agent exhibits a certain bias towards selecting some of the links – the bias is induced by the background information that the agent has about the network. Another distinction between a random walk and an informed search is the stochasticity of their link selection mechanism. By definition a random walk is a stochastic dynamical process, whereas in an informed search the link selection mechanism can be (i) either deterministic in cases where the agent e.g. acts greedy and always follows an optimal link (according to a given criterion), (ii) or stochastic in cases where the agent draws a link from a given probability distribution (induced by the background knowledge).

Various studies investigated the nature of the background knowledge in e.g. informed search in social networks. For example, in studies of human search behavior that have been inspired by the famous small-world experiment by Milgram [1] the background knowledge about the network has been represented as a hierarchy of network nodes [2], [3]. The distance of nodes in the hierarchy has been used to inform deterministic decentralized search. Another type of the background knowledge has been analyzed in [4]. In this study a deterministic search is informed by the node degree (the number of nodes adjacent to a given node), a quantity that is always available locally as compared to a node hierarchy which assumes a global knowledge about the network. In [5] the authors built on this previous work and designed a stochastic algorithm that takes into account the node degree, as well as the nodes' homophily, i.e. the tendency of nodes to connect to other similar nodes. The presented algorithm probabilistically weighted the influence of the node degree and the node homophily on the link selection probability. Moreover, the algorithm modeled a situation in which an agent does not have information on the node similarities, in which case the decision was based solely on the node degree.

Similarly to social networks search in information networks is always local, i.e. users have at their disposal only the links going out from their current Web page. Human search behavior in information networks has been investigated in the work of West and Leskovec [6]. They analyzed click paths of users playing a navigation game on Wikipedia and in their subsequent work [7] they compared decentralized search algorithms using various distance functions and benchmarked them against the human navigation paths. The authors found that automatic deterministic search strategies with a complete background knowledge typically outperform human information seeking.

One of the most interesting findings was that human navigation in information networks exhibits two phases: (i) Zoom-out phase in which the users strive to reach the network core, or a hub in the network core, e.g. an overview Wikipedia page with many links to different parts of the Wikipedia. In this first phase the humans tend to base their search decisions on the node degrees, i.e. they tend to select high-degree nodes. (ii) Zoom-in phase in which the users leave the core and close in on the topic of their interest, i.e. they dive in into a Wikipedia cluster (community) dealing with a given topic. In this second phase the humans base their decisions primarily on the similarity between network nodes such as textual similarity between Wikipedia nodes.

Thus, similarly to [5] where the authors engineered an algorithm that interchangeably bases its decisions on the node degree and the similarity between nodes West et al. [6], [7] found empirically that humans base their search decisions on

those two quantities – alternating between the node degree in the first phase and the node similarity in the second phase.

Although the previous research has identified the importance of informing search in social as well as in information networks with the node degree and the similarity between nodes not much work has been done on the analysis of various network internal similarity measures and their ability to capture those two network properties. For example, in those previous studies [5]–[7] the similarity has been calculated from the external information about the network nodes: in [5] the authors calculated standard vector-based similarity on the titles of the scientific articles in the citation networks and in [6], [7] the authors applied the same similarity measure but calculated it on the complete text of Wikipedia articles.

However, it is an interesting research question to analyze the properties of the similarity measures that are based on the internal network linking patterns since the presence of the external information is not always guaranteed and its quality may greatly vary from one network to another. Moreover, measures based on the linking patterns can in a general case capture both node properties: its degree and its similarity to other nodes in a single quantity. Therefore, it seems important to investigate if and how such measures are able to inform search in networks. Thereby, we can concentrate on the suitability of these measures to reflect e.g. the two phase process of the navigation in social and information networks, where in the first phase the node degree plays the most important role and in the second phase that role is taken by the node similarity.

In this paper, we turn our attention to the analysis of internal linking patterns to serve as a similarity measure for informing search in networks. To that end, we concentrate on simple algebraic measures that capture so-called regular equivalence and analyze the distribution of such similarities over a range of synthetic networks. In particular we concentrate on Katz similarity as an example of a measure of regular equivalence in networks. Thereby, we are interested how Katz similarity captures heterogeneity in node degrees and the existence of highly interlinked clusters of similar nodes (network communities).

The paper is organized as follows. In the next Section we give a short overview of the related work in this field. In Section III we describe our methodology and Section IV gives the details about our experiments with synthetic networks. In Section V we present the experimental results and shortly discuss them. Finally, we conclude the paper and provide directions for further research.

II. Related Work

Search has been an important concept in theoretical and empirical studies of networks. For example, in Web search the famous PageRank calculation [8], which determines the importance of Web pages is based on a random walk on the Web network. In a so-called Random Surfer model a hypothetical user navigates the Web by clicking randomly on the Web links. In the limit of large number of clicks the fraction of visits to each Web page gives its PageRank. Another example of an application of the random walk includes the problem of detecting communities (groups of highly connected and similar nodes) in networks. For example, it has been shown that designing minimal codes for encoding paths of a random walker in a network is equivalent to optimally partitioning of the network into communities [9]. Thus, we can design community detection algorithms that aim at optimally compressing a random walk paths in a network.

Research on decentralized search in social networks started with Milgram's seminal small world experiment [1] in which he aimed to study the connectedness of the US society. Milgram found that people in such a large social network are connected by short chains of acquaintances. Moreover, he found that people are able to navigate large social networks efficiently, i.e. that they are able to find those short chains even if they only possess the local knowledge of the network. In the subsequent research, Kleinberg analyzed the second result of the Milgram's experiment – the ability of humans to find a short path when there is such a path between two nodes [2], [10], [11]. Kleinberg concluded that social networks possess certain latent properties that humans are aware of. This background knowledge of network structure allows humans to find a short path between two arbitrary network nodes efficiently. Kleinberg has also investigated the nature of background knowledge that is required for efficient decentralized search algorithms. In other words: What structural properties do efficiently navigable networks possess? To that end, Kleinberg designed a number of network models such as the 2D-grid model [10], hierarchical model [11], and group model [11]. Independently, Watts [3] introduced the notion of social identity as a membership in a number of social groups organized in hierarchies and showed the existence of efficient decentralized search algorithms by simulation. Finally, Adamic [12] and others applied such a model to explain search in e.g. professional social networks.

All of these models have been based on the assumption that an agent navigating the network possesses a complete background knowledge on distances or similarities between pairs of network nodes. Contrary to that, in [4] the authors argued that distances between nodes are typically available only for some of the node pairs and that the assumption that an agent possesses a complete background knowledge is unrealistic. Thus, the authors analyzed an informed search algorithm that bases its decisions on the node degree (the number of nodes adjacent to a given node), a quantity that is always available locally. The algorithm performed better in comparison with a random walk but could not reach the performance of an agent having a complete background knowledge.

In an information network decentralized search can be used to model user navigation in an information network, e.g. the Web or Wikipedia [6]. In that model the user navigates from a Web page to a Web page by following her own intuition and her own knowledge about the topics of various Web pages.

III. METHODOLOGY

A. Node Similarity

A crucial element of an informed search is the notion of distance or similarity between the pairs of network nodes. Mathematically, we may represent the background knowledge with a distance or a similarity matrix. The rows and columns of the similarity matrix represent the network nodes and the elements of the matrix give the similarity score between corresponding network nodes. If the similarity function is symmetric we obtain a symmetric similarity matrix. Partial background knowledge may be easily represented by blank elements in the matrix for those node pairs where the agent's knowledge is missing. In this paper we denote the similarity matrix with $\boldsymbol{\sigma}$ – the element σ_{ij} of the matrix gives the similarity between nodes i and j from the network.

Different similarity measures have been previously discussed in the literature for various tasks. For example, [13] gives an overview and evaluation of various similarity measures for the link prediction task. In general, similarities in networks can be calculated from external or exogenous information about the networks (e.g. textual content in an information networks), or from internal or endogenous information (internal linking patterns). There are two possible similarity measures based on the network linking patterns: structural equivalence measures and regular equivalence measures. Structural equivalence is based on the simple idea that two nodes are similar if they share many of the same network neighbors, whereas regular equivalence between two nodes is given if they have neighbors who themselves are similar. Since absolute numbers of common neighbors are hard to interpret, this number is typically normalized. Different normalizing constants give rise to different similarity measures such as cosine similarity, Pearson correlation coefficients, or Euclidean distance ([14] gives a nice overview and comparison of various normalization methods). In contrast, regular equivalence between nodes is established when two nodes do not necessarily share same neighbors but rather they have neighbors who themselves are similar, or in a more simple case two network nodes i and j are similar if j is similar to other neighbors of i. In past, several simple algebraic measures have been developed in the past to measure regular equivalence [15], [16].

Previous research showed that node degree as well as node similarity play an important role in informed search in many networks. Therefore, we concentrate in this paper on the analysis of regular equivalence measures since structural equivalence can not capture both of these aspects in a single quantity. These measures assign a zero similarity to all nodes that do not have neighbors in common – they only capture local similarity between nodes but can not reflect any global or long range dependences between nodes. They are suitable to inform search if the agent is already in the proximity of its goal, but they can not lead that agent from a distant part of the network to the cluster or community where the target is situated. In other words, structural equivalence would not be able to instruct an agent in its zoom-out phase but only in its zoom-in phase. On the other hand, because regular equivalence measures are defined recursively over node neighbors, they are able to capture global node similarities that go beyond immediate node neighbors. They assign a non-zero similarity value to each node pair that exhibits a structural dependence in the network, e.g. the nodes are connected by at least one path in the network.

B. Katz Similarity

In this paper we concentrate on a particular measure of regular equivalence called Katz similarity [13]. Let us first introduce the basic quantities relevant for measuring this similarity. Let us denote the adjacency matrix of a graph $G(V, E)$ with $\boldsymbol{A}$:

$$A_{ij} = \begin{cases} 1 \text{ if nodes } i \text{ and } j \text{ are connected by a link} \\ 0 \text{ otherwise.} \end{cases} \quad (1)$$

We define Katz similarity recursively as [17]:

$$\sigma_{ij} = \alpha \sum_k A_{ik}\sigma_{kj} + \delta_{ij}, \quad (2)$$

where δ_{ij} is Kronecker delta ($\delta_{ij} = 1$ if and only if $i = j$, otherwise $\delta_{ij} = 0$).

The idea is that the term σ_{kj} is large if the neighbor k of i is similar to j. We then sum over all neighbors of i, which is ensured by the product $A_{ik}\sigma_{kj}$ since A_{ik} equals 1 only if i and k are neighbors. Finally, the term δ_{ij} assigns a high value of "self-similarity" of a node to itself. The constant α weights the influence of neighbors as compared to the influence of the self-similarity to the nodes similarity.

We can write the expression for the Katz similarity in the matrix form as:

$$\boldsymbol{\sigma} = \alpha \boldsymbol{A}\boldsymbol{\sigma} + \boldsymbol{I}, \quad (3)$$

where $\boldsymbol{I}$ is the identity matrix.

We can now solve this equation for $\boldsymbol{\sigma}$, which gives us the following:

$$\boldsymbol{\sigma} = (\boldsymbol{I} - \alpha \boldsymbol{A})^{-1}. \quad (4)$$

In case of large networks with e.g. millions of nodes, inverting a matrix is computationally expensive. Therefore, we may want to evaluate the Equation 3 recursively. We would typically start by assigning initial values for similarity to zero, e.g. $\boldsymbol{\sigma}^0 = 0$. We then obtain (for the first three iterations):

$$\boldsymbol{\sigma}^0 = 0, \boldsymbol{\sigma}^1 = \boldsymbol{I}, \boldsymbol{\sigma}^2 = \alpha \boldsymbol{A} + \boldsymbol{I}, \boldsymbol{\sigma}^3 = \alpha^2 \boldsymbol{A}^2 + \alpha \boldsymbol{A} + \boldsymbol{I}. \quad (5)$$

Thus, in the limit of large number of iterations m we have:

$$\boldsymbol{\sigma} = \sum_{m=0}^{\infty} \alpha^m \boldsymbol{A}^m. \quad (6)$$

The elements of matrix $\boldsymbol{A}^m$ at index i and j, i.e. $[\boldsymbol{A}^m]_{ij}$ give the total number of paths of length m between nodes i and j. The Katz similarity can be seen as a weighted count of all paths between the nodes i and j with paths of length r getting weight α^r. As long as $\alpha < 1$ longer paths obtain

less weight than shorter ones.Thus, two nodes will be similar if they are connected either by few short paths or if they are connected by many longer paths.

The inverse of $(\boldsymbol{I}-\alpha\boldsymbol{A})$ does not exist for $det(\boldsymbol{I}-\alpha\boldsymbol{A})=0$, which is exactly the characteristic equation for eigenvalues of $\boldsymbol{A}$, i.e. in the matrix form $\boldsymbol{\lambda}=\mathbf{1}/\boldsymbol{\alpha}$. Thus, the inverse does not exist whenever $1/\alpha$ equals λ_i. If we start with $\alpha=0$ and gradually increase it, we first hit the equality $1/\alpha=\lambda_i$ for the largest eigenvalue λ_1, and then again every time when we approach the next eigenvalue. Thus, if we pick an $\alpha<1/\lambda_1$ we ensure that the inverse exists (or that $\sum_{m=0}^{\infty}\alpha^m\boldsymbol{A}^m$ converges).

Therefore, we may set the value of parameter α for each network as:

$$\alpha=\frac{1}{\lambda_1}f, \tag{7}$$

where λ_1 is the largest eigenvalue of the adjacency matrix $\boldsymbol{A}$, and f is a parameter in the interval $[0.01, 0.99]$. We call f the α-parameter.

C. Synthetic networks

We carry our experiments on synthetic networks. The goal is to generate networks that structurally closely reflect a typical social or an information network. Although the real networks exhibit a wide range of interesting structural properties we concentrate on two which have been observed in the majority of networks from different domains: (i) degree heterogeneity (where we typically have only a few of high degree and mid degree nodes that keep the network connected), (ii) and modular or community structure (with various groups of nodes that are tightly connected with each other and only loosely connected to other communities – typically via high and mid degree nodes). Such networks reflect also typical search scenarios such as search for a similar node within the same community and search for a distant target node in another community. The first scenario is simpler and consists only of the zoom-in phase where an agent needs only to close in on the target node, whereas the second scenario is a more complicated one and contains both search phases: zoom-out phase where the agent exits the starting community and enters the network core and zoom-in phase where the agent enters the target community and finally closes in on the target node.

In our experiments we want to analyze if Katz similarity can capture those two phases in a search process. Thus, for a given target node i:

1) The most similar nodes to i should be the nodes from its community – this would instruct an agent that entered the i's community and is already in the proximity of its target to explore that community in its search for the target.
2) The nodes from the other communities should have low similarity to the given target node i – this would prohibit an agent to enter a distant community.
3) The nodes that keep the network connected such as high and mid degree nodes should have high to moderate similarity values to a given node i – this would guide an agent to leave a distant community in the zoom-out phase. Combined with the high similarity to the nodes from the own community it would further instruct the agent to enter the i's community and explore it until it reaches its goal.

Thus, we generate networks with injected communities and heterogeneous node degrees. A standard generative model for such networks are so-called stochastic block models, which we describe next.

D. Stochastic Block Models

Stochastic Block Models (SBM) are generative models for networks with clusters or communities [18]–[23]. They are simple extensions of the standard Erdős–Rényi [24] or Bernoulli graph model. In Bernoulli model any pair of nodes from the network is connected with a constant probability p. In addition, each link is independent on any other link in the network. This model generates networks with rather homogeneous (Poisson) degree distributions. In a stochastic block model nodes are grouped into k communities and the probability of a link between two nodes is conditioned on their membership in different communities. Thus, a stochastic block model is a mixture of k^2 Bernoulli models with different linking probabilities.

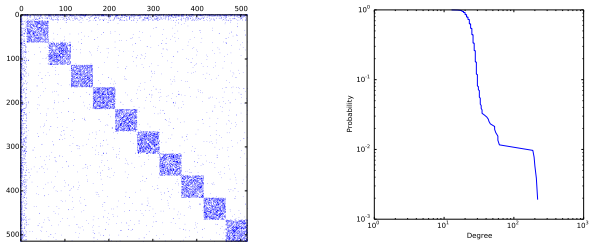
Mathematically, a stochastic block models is specified by: (i) the number of the communities k, (ii) a vector $\boldsymbol{s}$, which defines the membership of nodes in communities, and (iii) a $k\times k$ matrix $\boldsymbol{P}$, which defines probabilities of links between nodes from different groups.

The element P_{uv} of the matrix $\boldsymbol{P}$ defines the linking probability of nodes belonging to communities u and v. For the elements on the diagonal P_{uu} we obtain the linking probabilities within communities and the off-diagonal elements represent the probabilities of links that run between communities. To obtain a network with tightly connected communities with only a few links that connect the nodes from different communities we set higher probabilities in the diagonal elements P_{uu} and lower probabilities in the off-diagonal elements (see Equation 8).

To obtain a heterogeneous degree distribution we need a special community (or a couple of such communities) which are smaller than the other communities and which reverse the linking probabilities: the nodes from this community have low probability for connecting to other members of their community, but in turn they have high probability of connecting to all other communities. As a consequence these nodes obtain a larger number of links than average and thus represent high-degree or hubs in the network. The same modeling concept may be further applied to e.g. obtain mid-degree nodes, which also have a degree larger than average but still smaller than the hubs – the probabilities for connecting to the nodes from other communities should be larger than for typical nodes but smaller than the hub probabilities.

IV. Experiments

In our experiments we generate networks with 12 communities with one high-degree community and one mid-degree



(a) Adjacency matrix (b) Cumulative degree distribution

Fig. 1. **Left: Community structure.** The community structure is clearly visible in the visualization of the adjacency matrix. The thin blue line at the top and to the left of the matrix is the community of high-degree nodes. A blueish area beneath the top and to the right of the thin blue line is the mid-degree community. **Right: Heterogeneous degree distribution.** Cumulative degree distribution exhibits the existence of a small number of highly connected nodes (hubs).

community, which have 5 and 10 nodes respectively. All other 10 communities have 50 nodes. In the probability matrix for the high-degree community we assign a high probability of those nodes connecting to all other nodes from other communities. For mid-degree community we apply the same mechanism but assign smaller probabilities than for the high-degree community. All other 10 communities have a small probability of connecting to a different community (low off-diagonal entries) but a high probability of connecting to itself (high diagonal entries):

$$\boldsymbol{s} = \begin{pmatrix} 5 \\ 10 \\ 50 \\ 50 \\ \vdots \end{pmatrix} \boldsymbol{P} = \begin{pmatrix} 0.000 & 0.400 & 0.400 & 0.400 & \dots \\ 0.400 & 0.000 & 0.100 & 0.100 & \dots \\ 0.400 & 0.100 & 0.400 & 0.005 & \dots \\ 0.400 & 0.100 & 0.005 & 0.400 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{8}$$

A typical adjacency matrix and a cumulative degree distribution are shown in Figure 1.

Now, we are interested in the average similarity of nodes to:

1) the nodes from their own community (Own),
2) the nodes from the other communities (Other),
3) the nodes from a moderate size community of mid-degree nodes (Mid),
4) the nodes from a small size community of high-degree nodes (High).

We investigate how these average similarities change as a function of the α-parameter f from Equation 7. Thus, we iterate over the interval $[0.01, 0.99]$ in steps of 0.01 for a total of 99 different parameter settings. For each setting we generate $1,000$ random networks (for a total of $99,000$ generated networks for all parameter configurations) using a stochastic block model with the group vector $\boldsymbol{s}$ and the probability matrix $\boldsymbol{P}$ from Equation 8. For each parameter value we plot the average similarities and the standard deviation over $1,000$ generated networks.

V. Results and Discussion

Figure 2 shows the results of our experiments. For smaller values (e.g. f less than 0.1) of the α-parameter the most similar nodes to arbitrary nodes are, on average, high-degree nodes and the nodes from their own community. The mid-degree nodes have moderate average similarity and the nodes from communities other than the own community have a small average similarity. Thus, these distributions of similarity are suitable for informing search in networks. For example, these configurations can support zoom-out phase, where an agent starting in a distant community needs first to leave that community and access nodes that provide shortcuts to the target community. High-degree nodes connect different parts of the network and an agent needs to visit them first to close in on the target community. Since the high-degree nodes have a large average similarity to an arbitrary node in a network an agent following this similarity intuition would first visit one of the high-degree nodes. On the other hand, once when the agent reaches the target community the most similar nodes to the target node are the nodes from its own community – thus, an agent would enter the target community and remain within it to explore the community nodes and eventually reach its target.

For mid range of values of α-parameter (greater than 0.1 and less than 0.9) the high-degree nodes start to dominate other node types and become the most similar nodes to arbitrary nodes regardless of their community. Thus, an agent would correctly leave a distant community by visiting one of the high-degree nodes. However, the agent would remain in the network core indefinitely long. Even if the agent succeeds in reaching the target community it would in most cases try to reach the target node by leaving the target community and revisiting the high-degree nodes. The exploration of the target community would not take place at all.

For even larger values of the α-parameter (greater than 0.9) even the mid-degree nodes become more similar to arbitrary nodes than the nodes from their communities. Thus, an agent would in most cases traverse links connecting high-degree and mid-degree nodes hoping in the network core and failing even to reach the target community.

Summarizing, regular equivalence measures in general and Katz similarity in particular can provide the background knowledge and inform search in networks efficiently. They are able to capture both phases of the search process: exiting a distant community and entering the network core with many long range links to the target community, as well as the exploration of the target community. However, it is necessary to configure the parameters for calculating the regular equivalence measures appropriately. Our experiments show that the desired behavior of these measures can be only obtained by lower values of the control α-parameter, i.e. for values less than 0.1.

VI. Conclusions

In this paper we provide a short overview of various search process on structurally complex networks. We then analyze the

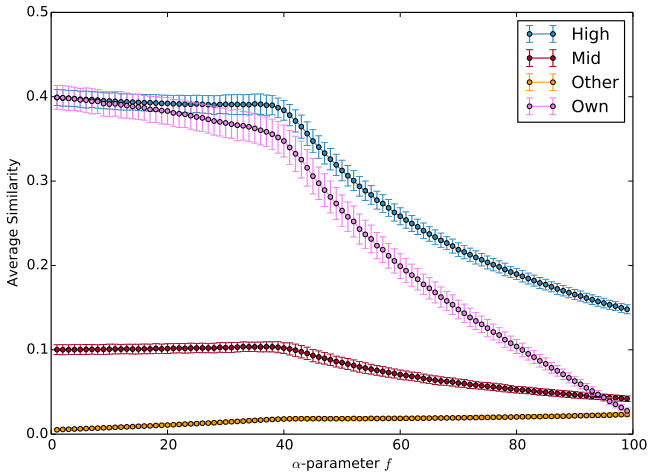


Fig. 2. **Average similarity as a function of α-parameter** For smaller values of the path weights the most similar nodes are those from the own community, closely followed by the hubs. The average similarity of the mid degrees is also significantly higher than the average similarity to nodes in the other communities. For large values of the path weights the similarity of high degree nodes dominates all other communities. The similarity to the nodes from the own community rapidly drops and for the values above 0.9 even the mid degree nodes become more similar on the average than the nodes from the own community. From the informed search point of view, the smaller values of α-parameter (i.e. values less than 0.1) are more suitable for providing guidance in both phases of search: zoom-out phase where high-degree nodes must be quickly reached, and the zoom-in phase where the nodes from the target communities need to be the most similar nodes.

suitability of regular equivalence measures to inform search in such networks. In particular, we analyze simple algebraic measures such as Katz similarity which capture several network structural properties such as node degree and node similarity into a single similarity score. Our experiments on synthetic networks with heterogeneous node degree distributions and injected community structure show that Katz similarity can efficiently inform search in networks. However, the parameters used to calculate the similarity need to be chosen with care and appropriately.

The main limitation of our work is the theoretical nature of the analysis. Specifically, we estimated the theoretical suitability of Katz similarity to inform search in networks. An experimental proof of concept would be also needed. For example, one can simulate an agent performing greedy search in networks by using the Katz similarity as the background knowledge and measure the agent's success rate and its search efficiency. Furthermore, the experiments should be repeated on empirical social and information networks that exhibit typical structural properties such as heterogeneous node degrees and community structure. However, we leave these experiments for the future work.

ACKNOWLEDGMENT

This research was in part funded by the FWF Austrian Science Fund research project "Navigability of Decentralized Information Networks" (P 24866-N15).

REFERENCES

[1] S. Milgram, "The small world problem," *Psychology Today*, vol. 1, pp. 60–67, 1967.
[2] J. M. Kleinberg, "Navigation in a small world," *Nature*, vol. 406, no. 6798, p. 845, August 2000.
[3] D. J. Watts, P. S. Dodds, and M. E. J. Newman, "Identity and search in social networks," *Science*, vol. 296, pp. 1302–1305, 2002.
[4] L. A. Adamic, R. M. Lukose, A. R. Puniyani, and B. A. Huberman, "Search in power-law networks," *Physical Review E*, vol. 64, no. 4, pp. 046 135 1–8, Sep 2001.
[5] O. z. Simsek and D. Jensen, "Decentralized search in networks using homophily and degree disparity," in *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, ser. IJCAI'05. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2005, pp. 304–310.
[6] R. West and J. Leskovec, "Human Wayfinding in Information Networks," in *Proceedings of the 21th international conference on World Wide Web - WWW '12*. New York, New York, USA: ACM Press, 2012.
[7] ——, "Automatic versus human navigation in information networks," in *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media, ICWSM*, 2012.
[8] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," in *Proceedings of the seventh international conference on World Wide Web 7*, ser. WWW7. Amsterdam, The Netherlands, The Netherlands: Elsevier Science Publishers B. V., 1998, pp. 107–117.
[9] M. Rosvall, D. Axelsson, and C. T. Bergstrom, "The map equation," *The European Physical Journal Special Topics*, vol. 178, no. 1, pp. 13–23, 2009.
[10] J. Kleinberg, "The small-world phenomenon: an algorithm perspective," in *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, ser. STOC '00. New York, NY, USA: ACM, 2000, pp. 163–170.
[11] ——, "Small-world phenomena and the dynamics of information," in *Advances in Neural Information Processing Systems (NIPS) 14*. Cambridge, MA, USA: MIT Press, 2001, pp. 431–438.
[12] L. Adamic and E. Adar, "How to search a social network," *Social Networks*, vol. 27, no. 3, pp. 187 – 203, 2005.
[13] D. Liben-Nowell and J. Kleinberg, "The link prediction problem for social networks," in *Proceedings of the twelfth international conference on Information and knowledge management*, ser. CIKM '03. New York, NY, USA: ACM, 2003, pp. 556–559.
[14] E. A. Leicht, P. Holme, and M. E. J. Newman, "Vertex similarity in networks," *Phys. Rev. E*, vol. 73, p. 026120, Feb 2006.
[15] V. D. Blondel, A. Gajardo, M. Heymans, P. Senellart, and P. V. Dooren, "A measure of similarity between graph vertices: Applications to synonym extraction and web searching," *SIAM Rev.*, vol. 46, no. 4, pp. 647–666, Apr. 2004.
[16] G. Jeh and J. Widom, "Simrank: A measure of structural-context similarity," in *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '02. New York, NY, USA: ACM, 2002, pp. 538–543.
[17] M. Newman, *Networks: An Introduction*. New York, NY, USA: Oxford University Press, Inc., 2010.
[18] P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: first steps," *Social Networks*, vol. 5, no. 2, pp. 109–137, 1983.
[19] K. Faust and S. Wasserman, "Blockmodels: Interpretation and evaluation," *Social Networks*, vol. 14, no. 1–2, pp. 5 – 61, 1992, special Issue on Blockmodels.
[20] C. J. Anderson, S. Wasserman, and K. Faust, "Building stochastic blockmodels," *Social Networks*, vol. 14, no. 1–2, pp. 137 – 161, 1992, special Issue on Blockmodels.
[21] T. A. Snijders and K. Nowicki, "Estimation and prediction for stochastic blockmodels for graphs with latent block structure," *Journal of Classification*, vol. 14, no. 1, pp. 75–100, 1997.
[22] Y. J. Wang and G. Y. Wong, "Stochastic blockmodels for directed graphs," *Journal of the American Statistical Association*, vol. 82, no. 397, pp. pp. 8–19, 1987.
[23] A. Goldenberg, A. X. Zheng, S. E. Fienberg, and E. M. Airoldi, "A survey of statistical network models," *Found. Trends Mach. Learn.*, vol. 2, no. 2, pp. 129–233, Feb. 2010.
[24] P. Erdos and A. Renyi, "On the evolution of random graphs," *Publ. Math. Inst. Hung. Acad. Sci*, vol. 5, pp. 17–61, 1960.