



BIROn - Birkbeck Institutional Research Online

De Meo, P. and Levene, Mark and Provetti, Alessandro (2019) Potential gain as a centrality measure. In: IEEE/WIC/ACM International Conference on Web Intelligence (WI '19), 14-17 October 2019, Thessaloniki, Greece.

Downloaded from: <https://eprints.bbk.ac.uk/id/eprint/31942/>

Usage Guidelines:

Please refer to usage guidelines at <https://eprints.bbk.ac.uk/policies.html> or alternatively contact lib-eprints@bbk.ac.uk.

Potential gain as a centrality measure

PASQUALE DE MEO, Dept. of Ancient and Modern Civilizations

University of Messina, Italy

MARK LEVENE and ALESSANDRO PROVETTI, Dept. of Computer Science

Birkbeck, University of London

Navigability is a distinctive features of graphs associated with artificial or natural systems whose primary goal is the transportation of information or goods. We say that a graph \mathcal{G} is navigable when an agent is able to efficiently reach any target node in \mathcal{G} by means of local routing decisions. In a social network navigability translates to the ability of reaching an individual through personal contacts. Graph navigability is well-studied, but a fundamental question is still open: why are some individuals more likely than others to be reached via short, friend-of-a-friend, communication chains? In this article we answer the question above by proposing a novel centrality metric called the *potential gain*, which, in an informal sense, quantifies the easiness at which a target node can be reached. We define two variants of the potential gain, called the *geometric* and the *exponential potential gain*, and present fast algorithms to compute them. The geometric and the potential gain are the first instances of a novel class of *composite centrality metrics*, i.e., centrality metrics which combine the *popularity* of a node in \mathcal{G} with its *similarity* to all other nodes. As shown in previous studies, popularity and similarity are two main criteria which regulate the way humans seek for information in large networks such as Wikipedia. We give a formal proof that the potential gain of a node is always equivalent to the product of its degree centrality (which captures popularity) and its Katz centrality (which captures similarity).

CCS Concepts: • **Theory of computation** → **Graph algorithms analysis**; • **Information systems** → **Web crawling**.

Additional Key Words and Phrases: Graph Navigability, Node Ranking in Graphs, Centrality

1 INTRODUCTION

Centrality metrics [28] provide a ubiquitous Network Science tool for the identification of the “important” nodes in a graph. They have been widely applied in a range of domains such as early detection of epidemic outbreaks [11], viral marketing [25], trust assessment in virtual communities [2], preventing catastrophic outage in power grids [3] and analyzing heterogeneous networks [1].

The notion of importance of a node can be defined in a number of ways [6–8, 30, 34]. Some centrality metrics define the importance of a node i in a graph \mathcal{G} as function of the distance of i to other nodes in \mathcal{G} : for instance, in *Degree Centrality*, the importance of i is defined as the number of the nodes which are adjacent to i , i.e. which are at distance one from i . Analogously, *Closeness Centrality* [30] classifies as important those nodes which are few hops away from any other node in \mathcal{G} .

Another class of centrality metrics looks at walk/path structures in \mathcal{G} : a walk is a sequence of adjacent nodes; its length is defined as the number of edges it contains; a path is a walk without repeated edges and the shortest path connecting two nodes is also called its *geodesic path*. For instance, the *Betweenness Centrality* [30] of i is the ratio of the number $g_{jl}(i)$ of geodesic paths from any node j to any node l which pass through the node i to the number g_{jl} of geodesic paths running from j to l and, thus, nodes with largest betweenness centrality scores are those which intercept most of the geodesic paths in \mathcal{G} .

A further popular metric is *Katz Centrality Score* [22], which is understood as the weighted number of walks terminating in i : here, the weighting factor is inversely related to walk length and, thus, long (resp., short) walks have a small (resp., large) weight.

For a suitable choice of the weighting factor, the Katz centrality score converges to the *Eigenvector Centrality* [5, 8] or the popular *PageRank* [8, 9].

To the best of our knowledge, however, there is no previous work in which the centrality of a node is closely related to the notion of *navigability*: roughly speaking, we say that \mathcal{G} is navigable if it is possible to successfully route a message to any node i in \mathcal{G} via a short chain of intermediary nodes, regardless of the node j which generates the message.

Navigability is one of the most important features for a broad range of natural and artificial systems which have the transportation of information (e.g. a computer network) or the trade of goods (e.g. a road network) as their primary purpose. In general, if the topology of the graph \mathcal{G} underlying the above mentioned systems would be perfectly specified, then any source node i could discover all shortest paths starting from (or terminating in) i and it could make use of the discovered paths to efficiently route messages.

In practice, nodes in \mathcal{G} are often able to efficiently route messages even if they do not have a global view of the topology of \mathcal{G} , and this has encouraged many researchers to seek a better understanding of why graphs arising in real applications are navigable. Early studies on graph navigability were inspired by the seminal work of Travers and Milgram [33] on the “small world” property.

In a celebrated experiment, random-chosen Nebraska residents were asked to send a booklet to a complete stranger in Boston. Selected individuals were required to forward the booklet to any of their acquaintances whom they deemed likely to know the recipient or at least might know people who did. In some cases, the booklet actually reached the target recipient by means, on average, of 5.2 intermediate contacts, thus suggesting an intriguing feature of human societies: in large, even planetary-scale, social networks, pairs of individuals are connected through *shorts chains of intermediaries* and ordinary people are able to uncover these chains [13, 17, 23, 26].

Several empirical studies have verified the small-world phenomenon in diverse domains such as metabolic and biological networks [21], the Web graph [10], collaboration networks among scientists [29] as well as social networks [13, 35].

So far, centrality metrics and navigability have been investigated in parallel, yet their research tracks are disconnected. Thus, an important (and still unanswered) direction of inquiry is the introduction of centrality metrics that are related to the *navigability* of a node, i.e., the ease at which it is possible to reach a target node i regardless of the node j chosen as source node.

In this article we tackle the questions above by extending previous work by Fenner *et al.* [16] to the realm of social networks. The main output of our research is an index, called the *potential gain*, which ranks nodes in a network on the basis of their ability to find a target.

The potential gain of a node i depends on the number of walks $w_k(j, i)$ of length k that connect i with any other node j . The underlying idea is that, for a fixed k , the larger $w_k(j, i)$, the higher the chance that j will reach i , regardless of the specific navigation strategy. In the computation of the potential gain, we take the small-world phenomenon as axiomatic: we consider an agent that starts from j and it looks for short walks to reach i .

We observe that the value a walk has for the agent will decrease with its length k and there is a threshold length beyond which the agent has to abandon that walk. To formalize the intuition above, we introduce a weighting factor $\phi(k)$ which monotonically decreases with k to penalize long walks.

We have developed two variants of the potential gain of [16], namely:

- the *geometric potential gain*, in which $\phi(k)$ decays as δ^k , where δ is a parameter ranging between 0 and the inverse of the spectral radius λ_1 of \mathcal{G}^1 , and
- the *exponential potential gain*, in which $\phi(k)$ decays in exponential fashion.

Both the geometric and exponential gain of i can be thought as the product of one index (Degree Centrality) related to the *popularity* of i and another (Katz Centrality score, for the geometric potential gain, and Communicability Index [4, 15] for the exponential potential gain) which reflects the degree of *similarity* of i with all other nodes in the network. In this sense, the geometric and the exponential potential gain are *composite centrality metrics*, i.e., they constitute a novel class of centrality metrics which combine popularity and similarity to rank nodes in graphs. The combination of popularity and similarity has proven to closely resemble the way humans navigate large social networks [31] or attempt to locate information in large information networks such as Wikipedia [18, 36, 37].

Our formalisation applies the *Neuman series expansion* [20] to efficiently but accurately approximate both the geometric and exponential gain. Both theoretical and experimental analysis show that our approach is appropriate for accurately computing the geometric and exponential potential gain in large real-life graphs consisting of millions of nodes and edges, even with modest hardware resources.

We validated our approach on three large datasets: FACEBOOK (a graph of friendships among Facebook users), DBLP (a graph describing scientific collaboration among researchers in Computer Science) and YOUTUBE (a graph mapping friendship relationships among YouTube users). The experimental results will be in the full version of this article.

2 BACKGROUND

In this section we introduce some basic terminology for graphs that will be largely used throughout this article.

Let a graph \mathcal{G} be an ordered pair $\mathcal{G} = \langle N, E \rangle$ where N is a set of *nodes*, here also called *vertices*, and $E = \{\langle i, j \rangle : i, j \in V\}$ is the set of *edges*. As usual, \mathcal{G} is *undirected* if edges are unordered pairs of nodes and *directed* otherwise. In this article we will consider only undirected graphs.

Also, let $n = |V|$ be the number of nodes, $m = |E|$ the number of edges of \mathcal{G} . For any given node i its neighborhood $\mathcal{N}(i)$ is the set of nodes directly connected to it; its *degree* d_i is the number of edges incident onto it, i.e., $d_i = |\mathcal{N}(i)|$.

A *walk* of length k (with k a non-negative integer) is a sequence of nodes $\langle i_0, i_1, \dots, i_k \rangle$ such that consecutive nodes are directly connected: $\langle i_\ell, i_{\ell+1} \rangle \in E$ for $\ell \in [0..k-1]$. Also, we use the term *path* for walks that do not have repeated vertices. A walk will be *closed* if it starts and ends at the same node.

We will represent graphs by their associated *adjacency matrix*, \mathbf{A} , defined as usual with $a_{ij} = 1$ if $\langle i, j \rangle \in E$ and 0 otherwise. Sometimes we may slightly simplify notation with $a_{ij} = \mathbf{A}_{ij}$.

The adjacency matrix provides a compact formalism to describe many graph properties: for instance, the matrix \mathbf{A}^2 where $a_{ij}^2 = \sum_{k=1}^n a_{ik}a_{kj}$, gives the number of walks of length two going from i to j . Inductively, for any positive integer m , the matrix \mathbf{A}^m will give the number of closed (resp., distinct) walks of length m between any two nodes i and j if $i = j$ (resp., if $i \neq j$) [12].

It is a well-know fact that the adjacency matrix of any undirected graph is *symmetric* and, hence, all its eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ are real. The largest eigenvalue λ_1 of \mathbf{A} is also called its *principal eigenvalue* or *spectral radius* of \mathcal{G} . Moreover, the corresponding eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ will form an orthonormal basis in \mathbb{R}^n [32]. Eigenpairs $\langle \lambda_i, \mathbf{v}_i \rangle$ are formed by the eigenvalue λ_i and the corresponding eigenvector \mathbf{v}_i .

¹The spectral radius of \mathcal{G} is defined as the largest eigenvalue of the adjacency matrix of \mathcal{G} .

3 A MODEL OF NETWORK NAVIGABILITY

In this section we introduce our new centrality metrics, called the geometric and exponential potential gain.

As we will see, they share a common physical interpretation which is based on the notion of *graph navigability*: roughly speaking, we say that a graph \mathcal{G} is navigable if, for any target node i in \mathcal{G} , it is possible to reach i via short paths/walks, independently of the node j (called *the source*) from which we choose to start exploring \mathcal{G} from.

In the light of previous research on graph navigability, we informally define the *navigability score* of a node i as a measure of the “easiness” with which it is possible to reach i independently of the source node j . In this way, the navigability score of a node can be interpreted as a *centrality metric*.

To define the navigability score we borrow some ideas from previous work by Fenner *et al.* [16], who formulated the problem of identifying a “good” page p from which a user should start exploring the Web. A page p is classified as a good starting point if it satisfies the following criteria: (1) *it is relevant*, i.e. the content of p closely matches user’s information goals, (2) *the page p is central*, i.e., the distance of p to other Web pages in the Web graph is as low as possible and (3) *the page p is connected*, in the sense that p is able to reach a maximum number of other pages via its outlinks.

A key difference between the approach of Fenner *et al.* [16] and the current one is that they defined the navigability score for i as the ability of i of acting as the *source node* for reaching all the other nodes. In our setting, instead, we think of the node i as the *target node* we wish to reach.

So, let us fix a source node j and a target node i and provide an estimate $\tau(j, i)$ of how “easy” it will be for i to be reached if we choose j as source node. Intuitively, the larger the number of walks from j to i , the easier it is for i to be reached from j ; in addition, we assume that the task of exploring a graph is costly and such cost increases as the length of the walks/paths we use for exploration purposes increases. Therefore, shorter walks should be preferred to longer ones.

By combining the requirements above, we obtain:

$$\tau(j, i) = \sum_{k=1}^{+\infty} \phi(k) \cdot w_k(j, i) \quad (1)$$

here $w_k(j, i)$ is the number of walks of length k going from j to i and the non-increasing function $\phi(k)$ acts as penalty for longer walks. If we sum over all possible source nodes j , we obtain a global centrality index $p(i)$ for i :

$$p(i) = \sum_{j \in \mathcal{N}} \tau(j, i). \quad (2)$$

In analogy to Fenner *et al.* [16, 27], we will call $p(i)$ the *potential gain* of i .

Depending on the choice of the penalty function $\phi(\cdot)$ we obtain two variants of the potential gain, namely the geometric and the exponential potential gain (see Section 3.1).

3.1 The geometric and exponential potential gain

Given the above specifications, we first define the potential gain in matrix notation. For the base case, consider walks of length $k=1$, i.e., direct connections. Only the neighbours of a node i will contribute to the potential gain of i , which leads to the trivial conclusion that, at $k = 1$, nodes with the largest degree are also those ones with the largest potential gain.

We define the vector \mathbf{p} such that $\mathbf{p}_i = p(i)$ for every node i :

$$\mathbf{p} = \phi(1) \cdot \mathbf{A} \times \mathbf{1}. \quad (3)$$

If we include walks of length two, then we have to consider the squared adjacency matrix \mathbf{A}^2 . So, we add a contribution $\phi(2) \cdot \mathbf{A}^2 \times \mathbf{1}$ to the potential gain.

By induction, nodes capable of reaching from i through walks of length up to k provide a contribution to the potential gain equal to $\phi(k) \cdot \mathbf{A}^k \times \mathbf{1}$. By summing over all possible values of k we get to the following expression for \mathbf{p} :

$$\begin{aligned} \mathbf{p} &= \phi(1)\mathbf{A} \times \mathbf{1} + \phi(2)\mathbf{A}^2 \times \mathbf{1} + \dots + \phi(k)\mathbf{A}^k \times \mathbf{1} + \dots \\ &= \sum_{k=1}^{+\infty} \left(\phi(k)\mathbf{A}^k \times \mathbf{1} \right) = \left(\sum_{k=1}^{+\infty} \phi(k)\mathbf{A}^k \right) \times \mathbf{1} \end{aligned}$$

To attenuate the effect of the walks' length, we will consider two weighting functions, namely:

(1) *Geometric*: $\phi(k) = \delta^{k-1}$ with $\delta \in (0, 1)$. So we define the *geometric potential gain*, \mathbf{g} :

$$\mathbf{g} = \left(\mathbf{A} + \delta\mathbf{A}^2 + \dots + \delta^{k-1}\mathbf{A}^k + \dots \right) \times \mathbf{1} \quad (4)$$

(2) *Exponential*: $\phi(k) = \frac{1}{(k-1)!}$. So we define the *exponential potential gain*, \mathbf{e} :

$$\mathbf{e} = \left(\mathbf{A} + \mathbf{A}^2 + \dots + \frac{1}{(k-1)!}\mathbf{A}^k + \dots \right) \times \mathbf{1} \quad (5)$$

4 POTENTIAL GAIN AS CENTRALITY

The geometric and the exponential potential gain introduced above yield a *ranking* of network nodes and, therefore, it is instructive to compare them with popular centrality metrics. Recall that we defined the *spectral radius* λ_1 of \mathbf{A} as the largest eigenvalue of \mathbf{A} .

As for the geometric potential gain, if we let $\delta < \lambda_1^{-1}$, the following expansion holds:

$$\begin{aligned} \mathbf{g} &= \left(\mathbf{A} + \delta\mathbf{A}^2 + \dots + \delta^{k-1}\mathbf{A}^k + \dots \right) \times \mathbf{1} \\ &= \mathbf{A} \times \left(\mathbf{I} + \delta\mathbf{A} + \dots + \delta^{k-1}\mathbf{A}^{k-1} + \dots \right) \times \mathbf{1} \\ &= \mathbf{A} \times (\mathbf{I} - \delta\mathbf{A})^{-1} \times \mathbf{1} \end{aligned}$$

in which we make use of the *Neuman series* [20]

$$\left(\mathbf{I} + \dots + \delta^{k-1}\mathbf{A}^{k-1} + \dots \right) = (\mathbf{I} - \delta\mathbf{A})^{-1}. \quad (6)$$

At this point, the term $(\mathbf{I} - \delta\mathbf{A})^{-1} \times \mathbf{1}$ is exactly the *Katz centrality score* [22, 24], a popular centrality metric that defines the importance of a node as a function of its similarity with other nodes in \mathcal{G} . Hence, we can say that the geometric potential gain combines two kind of contributions: *popularity*, as captured by node degree, and *similarity* as captured by Katz's similarity score.

It is also instructive to consider what happens for extreme values of δ : if $\delta \rightarrow 0$, then the geometric potential gain tends to $\mathbf{A} \times \mathbf{1}$, i.e., it coincides with degree. In contrast, if $\delta \rightarrow \frac{1}{\lambda_1}$, then the Katz centrality score converges to *eigenvector centrality* [5], another popular metric adopted in Network Science. Boldi et al. [6–8, 34] show that the Katz Centrality score is also strictly related to the PageRank. More specifically, the PageRank vector \mathbf{p} coincides with the Katz Centrality

score provided that the adjacency matrix \mathbf{A} is replaced by its row-normalized version $\bar{\mathbf{A}}$:

$$\mathbf{p} = (1 - \alpha) \sum_{k=0}^{+\infty} \alpha^k \bar{\mathbf{A}}^k \times \mathbf{1} \quad (7)$$

Here, the parameter α is the so-called PageRank *damping factor*. Let us now concentrate on the exponential potential gain. We rewrite Equation 5 as follows:

$$\begin{aligned} \mathbf{e} &= \left(\mathbf{A} + \mathbf{A}^2 + \dots + \frac{1}{(k-1)!} \mathbf{A}^k + \dots \right) \times \mathbf{1} \\ &= \mathbf{A} \times \left(\mathbf{I} + \mathbf{A} + \dots + \frac{1}{k!} \mathbf{A}^k + \dots \right) \times \mathbf{1} \\ &= \mathbf{A} \times \exp(\mathbf{A}) \times \mathbf{1} \end{aligned}$$

where $\exp(\mathbf{A}) = \sum_{k=0}^{+\infty} \frac{1}{k!} \mathbf{A}^k$ is the exponential of \mathbf{A} [19].

The exponential of a matrix has been used to introduce other centrality scores such as *communicability* or *subgraph centrality* [5, 14].

Specifically, $\exp(\mathbf{A})_{ij}$ measures how easy is to send a unit of flow from a node i to a node j and vice versa. Such a parameter is known as *communicability* and it can be regarded as a measure of similarity between a pair of nodes. Communicability has been successfully used to discover communities in networks [14]. The product $\exp(\mathbf{A}) \times \mathbf{1}$ yields a centrality metric which defines the importance of a node as function of its ability to communicate with all other nodes in the network. In turn, the diagonal entry $\exp(\mathbf{A})_{ii}$ of the matrix exponential defines a further centrality metric called *subgraph centrality* [15]. As a result of the rewriting above, we clearly see exponential potential gain as dependent on two factors: popularity of i (i.e., its degree) and similarity of i with all other nodes in the network.

The computation of the geometric (resp., exponential) potential gain for *all nodes* in \mathcal{G} needs the specification of the full adjacency matrix \mathbf{A} ; in this sense, the geometric and the exponential potential gain should be considered as *global centrality metrics*, on par with the Katz centrality score and Subgraph centrality.

5 CONCLUSIONS

We have introduced the potential gain, an index to rank nodes in graphs that captures the ability of a node to act as a target point for navigation within the network. We have defined two variants of the potential gain, the geometric and exponential potential gain. We then proposed two iterative algorithms that compute the geometric and exponential potential gain and proved their convergence. We evaluated the scalability of our algorithms on three real large datasets.

We have discovered connections between the geometric potential gain and other, well-known, centrality metrics; GPG provides a new, mixed global-local centrality measure. Indeed, the PG as a centrality index has several merits:

- it unifies Katz and Communicability into a single framework;
- in its definition in terms of the PG it allows us to provide novel and efficient approximations of these indices;
- it provides an instance of a novel class of composite indices, in this case DC*Katz, and
- as each vertex has clear visibility of its neighbours, the realisation that PG is a combination of local (Degree) and global (Katz) centrality makes complete sense, in our opinion.

It is also possible that these results will open the door to a new interpretation of social phenomena related to Travers-Milgram's "small world" experiment [33].

One question that could be discussed at this point is which of the two new measures could be considered the best analysis tool large networks. Early experimental results indicate different rates of convergence but no clear “winner.”

From a computational standpoint, the geometric potential gain is clearly superior. So, for the analysis of very large networks and/or modest hardware resources it is the navigability score of choice. One practical difference however remains. The exponential potential gain is parameter-free and can be applied directly. On the other hand, the geometric potential gain is parametric in δ thus it requires a careful tuning of the algorithm.

Another topic for future work is investigation on the relationship between network robustness and network navigability. To this end, we intend to design an experiment in which graph nodes are ranked on the basis of their geometric/exponential potential gain and then are progressively removed from the graph. Basic properties about graph topology, such as the number and size of connected components shall be re-evaluated upon node deletion. We also plan to study how adding edges can increase the geometric/exponential potential gain of a target group of nodes.

REFERENCES

- [1] S. Agreste, P. De Meo, E. Ferrara, S. Piccolo, and A. Provetti. 2015. Analysis of a Heterogeneous Social Network of Humans and Cultural Objects. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 45, 4 (2015), 559–570.
- [2] S. Agreste, P. De Meo, E. Ferrara, S. Piccolo, and A. Provetti. 2015. Trust Networks: Topology, Dynamics, and Measurements. *IEEE Internet Computing* 19, 6 (2015), 26–35. <https://doi.org/10.1109/MIC.2015.93>
- [3] R. Alber, I. Albert, and G.L. Nakarado. 2004. Structural vulnerability of the North American power grid. *Physical review E* 69, 2 (2004), 025103.
- [4] M. Benzi and C. Klymko. 2013. Total communicability as a centrality measure. *Journal of Complex Networks* 1, 2 (2013), 124–149.
- [5] M. Benzi and C. Klymko. 2015. On the limiting behavior of parameter-dependent network centrality measures. *SIAM J. Matrix Anal. Appl.* 36, 2 (2015), 686–706.
- [6] Paolo Boldi. 2015. Large-scale Network Analytics: Diffusion-based Computation of Distances and Geometric Centralities. In *Proceedings of the 24th International Conference on World Wide Web Companion, WWW 2015, Florence, Italy, May 18-22, 2015 - Companion Volume*, Aldo Gangemi, Stefano Leonardi, and Alessandro Panconesi (Eds.). ACM, 1313. <https://doi.org/10.1145/2740908.2741703>
- [7] Paolo Boldi, Alessandro Luongo, and Sebastiano Vigna. 2017. Rank monotonicity in centrality measures. *Network Science* 5, 4 (2017), 529–550. <https://doi.org/10.1017/nws.2017.21>
- [8] Paolo Boldi and Sebastiano Vigna. 2014. Axioms for Centrality. *Internet Mathematics* 10, 3-4 (2014), 222–262. <https://doi.org/10.1080/15427951.2013.865686>
- [9] S. Brin and L. Page. 1998. The anatomy of a large-scale hypertextual Web search engine. *Computer networks and ISDN systems* 30, 1-7 (1998), 107–117.
- [10] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. 2000. Graph structure in the Web. *Computer Networks* 33, 1-6 (2000), 309–320.
- [11] F. Chung, P. Horn, and A. Tsias. 2009. Distributing antidote using Pagerank vectors. *Internet Mathematics* 6, 2 (2009), 237–254.
- [12] D. Cvetkovic, P. Rowlinson, and S. Simic. 1997. *Eigenspaces of graphs*. Cambridge University Press.
- [13] P.S. Dodds, R. Muhamad, and D.J. Watts. 2003. An experimental study of search in global social networks. *Science* 301, 5634 (2003), 827–829.
- [14] E. Estrada, N. Hatano, and M. Benzi. 2012. The physics of communicability in complex networks. *Physics reports* 514, 3 (2012), 89–119.
- [15] E. Estrada and J.A. Rodriguez-Velazquez. 2005. Subgraph centrality in complex networks. *Physical Review E* 71, 5 (2005), 056103.
- [16] T. Fenner, M. Levene, and G. Loizou. 2008. Modelling the navigation potential of a Web page. *Theoretical Computer Science* 396, 1-3 (2008), 88–96.
- [17] S. Goel, R. Muhamad, and D.J. Watts. 2009. Social search in “Small-World” experiments. In *Proc. of the International Conference on World Wide Web (WWW 2009)*. Madrid, Spain, 701–710.
- [18] D. Helic, M. Strohmaier, M. Granitzer, and R. Scherer. 2013. Models of human navigation in information networks based on decentralized search. In *Proc. of the ACM conference on Hypertext and Social Media*. ACM, Paris, France, 89–98.
- [19] N.J. Higham. 2008. *Functions of matrices: theory and computation*. Vol. 104. SIAM.
- [20] R. Horn and C. Johnson. 2013. *Matrix analysis* (2 ed.). Cambridge Univ. Press.
- [21] H. Jeong, B. Tombor, R. Albert, Z. Oltvai, and A. Barabasi. 2000. The large-scale organization of metabolic networks. *Nature* 407, 6804 (2000), 651.
- [22] L. Katz. 1953. A new status index derived from sociometric analysis. *Psychometrika* 18, 1 (1953), 39–43.
- [23] J. Kleinberg. 2000. The small-world phenomenon: An algorithmic perspective. In *Proc. of the ACM symposium on Theory of computing (STOC 2000)*. ACM, 163–170.
- [24] E. Leicht, P. Holme, and M. Newman. 2006. Vertex similarity in networks. *Physical Review E* 73, 2 (2006), 026120.
- [25] J. Leskovec, L. Adamic, and B. Huberman. 2007. The dynamics of viral marketing. *ACM Transactions on the Web* 1, 1 (2007), 5.

- [26] J. Leskovec and E. Horvitz. 2008. Planetary-scale views on a large instant-messaging network. In *Proc. of the International Conference on World Wide Web, WWW 2008*. Beijing, China, 915–924.
- [27] M. Levene and R. Wheeldon. 2004. Navigating the World Wide Web. In *Web Dynamics*, M. Levene and A. Pouloussilis (Eds.). Springer, 117–151.
- [28] L. Lü, D. Chen, X. Ren, Q. Zhang, Y. Zhang, and T. Zhou. 2016. Vital nodes identification in complex networks. *Physics Reports* 650 (2016), 1–63.
- [29] M. Newman. 2001. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences* 98, 2 (2001), 404–409.
- [30] M. Newman. 2010. *Networks: an introduction*. Oxford University Press.
- [31] O. Simsek and D. Jensen. 2008. Navigating networks by using homophily and degree. *Proceedings of the National Academy of Sciences* 105, 35 (2008), 12758–12762.
- [32] G. Strang. 1993. *Introduction to linear algebra*. Vol. 3. Wellesley-Cambridge Press Wellesley, MA.
- [33] J. Travers and S. Milgram. 1967. The small world problem. *Psychology Today* 1, 1 (1967), 61–67.
- [34] Sebastiano Vigna. 2016. Spectral ranking. *Network Science* 4, 4 (2016), 433–445. <https://doi.org/10.1017/nws.2016.21>
- [35] D. Watts and S. Strogatz. 1998. Collective dynamics of small-world networks. *Nature* 393, 6684 (1998), 440.
- [36] R. West and J. Leskovec. 2012. Automatic Versus Human Navigation in Information Networks.. In *Proc. of the International Conference on Weblogs and Social Media, (ISWMC 2012)*. Dublin, Ireland.
- [37] R. West, J. Pineau, and D. Precup. 2009. Wikispedia: An Online Game for Inferring Semantic Distances between Concepts. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI 2009)*. Pasadena, California, USA, 1598–1603.