Ranking the Relationship on Wikipedia through Strength Determination

S.P.Angaiyarkanny[1], G.Sharmila[2]

Department of Information Technology, Syed Ammal Engineering College, Ramanathapuram, Tamilnadu, India[1]

Department of Computer Science and Engineering, Syed Ammal Engineering College, Ramanathapuram, Tamilnadu, India[2]

ABSTRACT— Relationships are there between objects in Wikipedia. This paper's intention is to measure the relationship between objects in Wikipedia and to ranking the relations based on their strength. Two sorts of relationships between two objects exist: in Wikipedia, an explicit relationship is pictured by one link between two pages for the objects; an implicit relationship is pictured by a link structure containing the two pages. A number of the antecedent planned ways for measurement of relationships are cohesion-based ways that underrate objects having high degrees, although such objects can be essential in constituting relationships in Wikipedia. The conflicting ways are insufficient for measurement of implicit relationships since they employ only one or two of the subsequent three necessary factors: distance, connectivity, and co citation. Here proposing a brand new methodology employing a generalized flow that reflects all the three factors and doesn't undervalue objects having high degree. Ensure through experiments that this methodology will give the strength of a relationship a lot of fittingly than these antecedent planned ways do. Another outstanding facet of this methodology is mining elucidatory objects, that is, objects that must constitute a relationship. Mining elucidatory objects would open completely unique thanks to deeply perceive a relationship.

KEYWORDS— Measurement, Ranking, Relationship, Generalized flow

I.   INTRODUCTION

Wikipedia is habitually a better option for a user to achieve knowledge of a single object than typical search engines. A user also might long to discover a relationship among two objects.

A relationship is an association between two or more people that may range in duration from brief to enduring. Word Association is a common word game involving an exchange of words that are associated together. This association may be based on regular business interactions, or some other type of social commitment. Interpersonal relationships are formed in the context of social. Interpersonal relationships are dynamic systems that change continuously during their existence.

A innovative method for measuring a relationship on Wikipedia by reflecting all the three concepts: distance, connectivity, and co citation. Here relationships are measured rather than the similarities. Assignment of the gain to each edge is important for measuring a relationship using a generalized maximum flow. It is established through experiments that the gain function is adequate to measure relationships appropriately. The strength of the relationship between a source object and each of its destinations objects, and rank the destination objects by the strength. The connection and production of other words in response to a given word is done spontaneously as a game, creative technique, or in a psychiatric evaluation.

The link analysis is a data-analysis technique used to evaluate relationships (connections) between nodes. Relationships may be identified among various types of nodes. Data gathering and processing requires access to data and has several inherent issues, including information overload and data errors. Once data is collected, it must be transformed into a format that can be effectively used by both human and computer analysers. The user can recognize an explicit relationship between two objects easily by reading the pages for the two objects in Wikipedia. By contrast, it

is tricky for the user to determine an implicit relationship and elucidatory objects without investigating a number of pages and links. So, it is an appealing problem to measure and explain the strength of an implicit relationship between two objects in Wikipedia.

A Wikipedia is a type of content management system, it differs from a blog or most other such systems in that the content is created without any defined owner or leader, and Wikipedia have little implicit structure, allowing structure to emerge according to the needs of the users. The encyclopaedia project Wikipedia is the most famous wiki on the public web, but there are many sites running many different kinds of wiki software. Wiki promotes meaningful topic associations between different pages by making page link creation almost intuitively easy and showing whether an intended target page exists or not. Wiki is not a carefully crafted site for casual visitors. Instead, it seeks to involve the visitor in an ongoing process of creation and collaboration that constantly changes the Web site landscape.

The nature of mining processes creates a potential negative impact on the environment both during the mining operations and for years after the mine is closed. This impact has led to most of the world's nations adopting regulations to moderate the negative effects of mining operations. Safety has long been a concern as well, and modern practices have improved safety in mines significantly. The data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection) and dependencies (association rule mining). This usually involves using database techniques such as spatial indices. These patterns can then be seen as a kind of summary of the input data, and may be used in further analysis or, for example, in machine learning and predictive analytics. For example, the data mining step might identify multiple groups in the data, which can then be used to obtain more accurate prediction results by a decision support system. Neither the data collection, data preparation, nor the result interpretation and reporting are part of the data mining step, but do belong to the overall knowledge discovery process as additional steps.

Data mining uses information from past data to analyse the outcome of a particular problem or situation that may arise. Data mining works to analyse data stored in data warehouses that are used to store that data that is being analysed. That particular data may come from all parts of business, from the production to the management. Managers also use data mining to decide upon marketing strategies for their product. They can use data to compare and contrast among competitors. Data mining interprets its data into real time analysis that can be used to increase sales, promote new product, or delete product that is not value-added to the company.

Text analysis involves information retrieval, lexical analysis to study word frequency distributions, pattern recognition, tagging/annotation, information extraction, data mining techniques including link and association analysis, visualization, and predictive analytics. The overarching goal is, essentially, to turn text into data for analysis, via application of natural language processing (NLP) and analytical methods. A typical application is to scan a set of documents written in a natural language and either model the document set for predictive classification purposes or populate a database or search index with the information extracted.

In Wikipedia, the knowledge of an object is gathered in a single page updated constantly by a number of volunteers. Wikipedia also covers objects in a number of categories, such as people, science, geography, politic, and history. Therefore, searching Wikipedia is usually a better choice for a user to obtain knowledge of a single object than typical search engines. A user also might desire to discover a relationship between two objects. Classic keyword search engines can neither measure nor explain the strength of a relationship. The main issue for measuring relationships arises from the fact that two kinds of relationships exist:"explicit relationships" and "implicit relationships."

This paper proceeds as follows. Section II describes the related work and the global operation of system. The different leveraged metrics and components of the system are presented in Section III. Section IV presents components. Section V explains the conclusion and future works.

II.   RELATED WORK

To rank the relationships between two objects in Wikipedia, measure the strength of the link among two pages that constitute the relationship. Existing system measures the relationship based on distance, connectivity and co citation.

A. Dataset

Dataset in Wikipedia includes objects that have relationships between each other. The existing method implies how to measure the strength of relationships among pages.

1)Loading a dataset

A data set is a collection of data, it lists values for each of the variables, such as height and weight of an object. The query used to generate a particular data set from the selected connection or flat file profile. Multiple data set definitions can be created for the same profile in order to generate different data set instances. To improve classification accuracy, insignificant parameters and patient data could be deleted from the data set.

The schema of a Data Set can be defined programmatically, created by the Fill or Fill Schema methods of a Data Adapter, or loaded from an XML document. To load Data Set schema information from an XML document, use either the Read Xml Schema or the Infer Xml Schema method of the Data Set can be used.

Read Xml Schema allows one to load or infer Data Set schema information from the document containing XML Schema definition language (XSD) schema, or an XML document with inline XML Schema. Infer Xml Schema allows to infer the schema from the XML document while ignoring certain XML namespaces that is specified.

2)Data Preprocessing

Data pre-processing is an important step in the data mining process. Data-gathering methods are often loosely controlled, resulting in out-of-range values. Analyzing data that has not been carefully screened for such problems can produce misleading results. Thus, the representation and quality of data is first and foremost before running an analysis.

If there is much irrelevant and redundant information present or noisy and unreliable data, then knowledge discovery during the training phase is more difficult. Data preparation and filtering steps can take considerable amount of processing time. Data pre-processing includes cleaning, normalization, transformation, feature extraction and selection, etc.

The product of data pre-processing is the final training set. The lacking of attribute values, lacking certain attributes of interest, or containing only aggregate data, reduce the volume but producing the same or similar analytical results.

B. Generalized Maximum Flow-based Method

The generalized maximum flow problem is identical to the classical maximum flow problem except that every edge e has a gain $\gamma(e)>0$; the value of a flow sent along edge e is multiplied by $\gamma(e)>0$. Let $f(e)\geq 0$ be the flow f on edge e, and $\mu(e) \geq 0$ be the capacity of edge e. The capacity constraint $f(e)\leq \mu(e)$ must hold for every edge e. The goal of the problem is to send a flow emanating from the source vertex s into the destination vertex t to the greatest extent possible, subject to the capacity constraints. Let generalized network G=(V,E,s,t μ, γ) be information network (V.E) with the source s$\acute{e}$V , the destination t$\acute{e}$V , the capacity μ, and the gain γ. Fig. 4 depicts an example of a generalized maximum flow on a generalized network.

Here proposing a new method for measuring the strength of a relationship using the generalized maximum flow. The value of flow f is defined as the total amount of f arriving at destination t. To measure the strength of a relationship from object s to object t, the value of a generalized maximum flow emanating from s as the source into t as the destination is used; a larger value signifies a stronger relationship. The vertices in the paths composing the generalized maximum flow as the objects constituting the relationship are regarded.
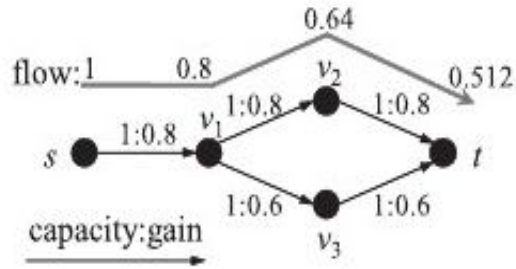
Fig. 4. Generalized maximum flow

1) Distance

In the methods based on distance, a shorter path represents a stronger relationship. For this method, $\gamma(e)<1$ for every edge e; then a flow considerably decreases along a long path. A short path usually contributes to the generalized maximum flow by a greater amount than a long path does. Therefore, a shorter path means a stronger relationship in the method.

2) Connectivity

In methods based on connectivity, a strong relationship is represented by many vertex disjoint paths from the source to the destination. The number of vertex disjoint paths can be computed by solving a classical maximum flow problem. The generalized maximum flow problem is a natural extension of the classical maximum flow problem. Therefore, it also can be used to estimate the connectivity.

3) Co citation

A flow emanates from the source into the destination, and therefore the flow seldom uses an edge whose direction is opposite that from the source to the destination. On the other hand, Use of both directions is required to estimate the co citation of two objects. Consider the relationship between two objects s and t in the network presented in Fig. 5a. Object u is co cited by s and t.
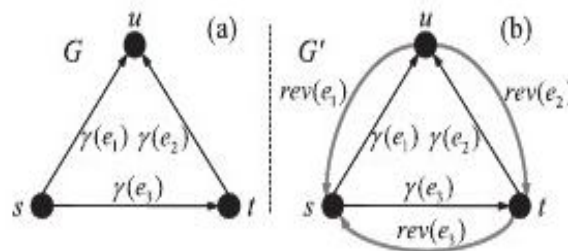

Fig. 5. A doubled network.

This co-citation is represented by two edges (s,u) and (t,u). However, unable to send a flow from s to t along the two edges, unless reverse the direction of the edge (t,u) to (u.t). Therefore, construct a doubled network by adding to every original edge in G a reversed edge whose direction is opposite to the original one. For example, Fig. 5b depicts the doubled network for the network presented in Fig. 5a.

III.   METHODS TO MEASURE THE RELATIONSHIP

A. Gain Function

In order to determine the gain function, first consider what kinds of explicit relationships are important in constituting an implicit relationship. In a generalized max-flow problem, a path composed of edges with large gain scan contributes to the value of a flow. To realize such a gain assignment, construct groups of objects in Wikipedia.

Categories cannot be used as groups directly because the category structure of Wikipedia is too fractionalized. Mutation can result in several different types of change in sequences, Mutations in genes can either have no effect, alter the product of a gene, or prevent the gene from functioning properly or completely.

B. Cycle-Free Effective Conductance (CFEC)

The cycle-free escape probability from s to t is the probability that a random walk originating at s will reach t without visiting any node more than once. The transition from one state to another does not depend on the previous state. The transition could be to the same state also. In network, proximity is the infinite number of attempts that is made to reach from starting node to end node.
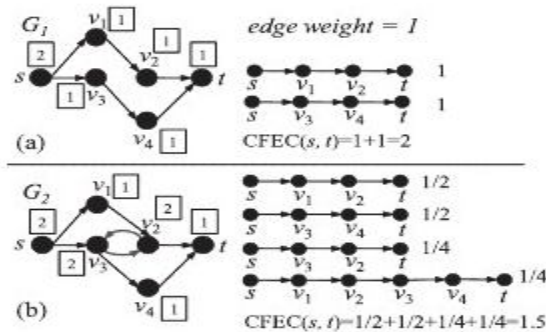

Fig. 1. CFEC on two networks

CFEC proximity allows to readily compute proximity graphs, which are small portions of the network that are aimed at capturing a related proximity value. It is extension of connection graph which is capable of presenting compact relationship between objects of a network. It can deduce relationship between more than two endpoints, the flexibility to handle edge direction, and the fact that they are obtained by solving an intuitively tunable optimization problem.

IV.   RANKING

A ranking is a relationship between a set of items such that, for any two items, the first is either 'ranked higher than', 'ranked lower than' or 'ranked equal to' the second. By reducing detailed measures to a sequence of ordinal numbers, rankings make it possible to evaluate complex information according to certain criteria. t is not

always possible to assign rankings uniquely. A common shorthand way to distinguish these ranking strategies is by the ranking numbers.

In competition ranking, items that compare equal receive the same ranking number, and then a gap is left in the ranking numbers. The number of ranking numbers that are left out in this gap is one less than the number of items that compared equal. The assignment of distinct ordinal numbers to items that compare equal can be done at random, or arbitrarily, but it is generally preferable to use a system that is arbitrary but consistent, as this gives stable results if the ranking is done multiple times. Query-independent methods attempt to measure the estimated importance of a page, independent of any consideration of how well it matches the specific query.

Source	Destinations	Human	Ours 3 hop	GSD	PFIBF 2 hop	CFEC 3 hop k=1000				THT d1 3 hop L_{max}=3
						ol	og	dl	dg	
Richard Nixon	Henry Kissinger	1 (8.5)	1 (2.31)	1 (0.26)	1 (7.94)	1 (1.24)	1 (0.91)	1 (1.89)	1 (1.12)	1 (2.98712)
	Zhou Enlai	2 (5.8)	2 (1.27)	2 (0.34)	2 (3.19)	2 (1.06)	2 (0.82)	2 (1.40)	2 (0.91)	3 (2.99098)
	Nguyen Van Thieu	3 (3.4)	3 (1.06)	2 (0.34)	3 (1.80)	3 (1.04)	3 (0.81)	3 (1.15)	3 (0.85)	4 (2.99173)
	Wallis Simpson	4 (2.0)	4 (0.80)	4 (0.49)	4 (0.45)	4 (1.00)	4 (0.80)	4 (1.02)	4 (0.81)	2 (2.98729)
Nobutaka Machimura	Yasuo Fukuda	1 (8.4)	1 (1.67)	1 (0.19)	1 (9.39)	1 (1.38)	1 (0.96)	1 (1.57)	1 (1.01)	1 (2.97889)
	Condoleezza Rice	2 (5.3)	2 (0.82)	2 (0.41)	3 (0.75)	3 (0.01)	3 (0.00)	2 (1.04)	2 (0.81)	2 (2.98354)
	George W. Bush	3 (4.1)	3 (0.64)	4 (0.56)	2 (1.14)	2 (0.02)	2 (0.01)	3 (0.09)	3 (0.03)	3 (2.99704)
	Hillary Clinton	4 (2.6)	4 (0.61)	3 (0.48)	4 (0.27)	4 (0.00)	3 (0.00)	4 (0.02)	4 (0.01)	4 (2.99886)
Donald Henry Rumsfeld	Dick Cheney	1 (7.7)	1 (2.05)	1 (0.17)	2 (3.38)	2 (1.08)	2 (0.84)	2 (1.25)	2 (0.90)	1 (2.96996)
	Condoleezza Rice	2 (6.9)	2 (1.47)	2 (0.22)	3 (2.58)	4 (0.02)	4 (0.01)	3 (0.23)	3 (0.09)	3 (2.98412)
	Ronald Reagan	3 (5.5)	3 (1.07)	3 (0.35)	1 (3.47)	1 (1.20)	1 (0.89)	1 (1.35)	1 (0.96)	2 (2.97003)
	Junichiro Koizumi	4 (3.8)	4 (0.46)	4 (0.53)	4 (1.63)	3 (0.06)	3 (0.02)	4 (0.10)	4 (0.03)	4 (2.99659)
Junichiro Koizumi	Shinzo Abe	1 (9.1)	1 (5.30)	1 (0.18)	1 (29.6)	1 (1.97)	1 (1.14)	1 (3.72)	1 (1.72)	1 (2.98931)
	Donald Rumsfeld	2 (5.3)	2 (1.99)	2 (0.53)	2 (2.32)	3 (0.12)	3 (0.04)	4 (0.098)	3 (0.03)	3 (2.99916)
	Wen Jiabao	3 (4.5)	4 (1.66)	2 (0.53)	4 (2.00)	2 (1.03)	2 (0.81)	2 (1.14)	2 (0.84)	2 (2.99666)
	Condoleezza Rice	4 (4.1)	3 (1.83)	4 (0.55)	3 (2.17)	4 (0.06)	4 (0.01)	3 (0.103)	4 (0.03)	4 (2.99948)
Bill Clinton	Hillary Clinton	1 (9.5)	1 (2.68)	1 (0.27)	1 (7.59)	1 (1.36)	1 (0.95)	1 (2.01)	1 (1.21)	1 (2.98550)
	Keizo Obuchi	2 (4.7)	4 (1.08)	3 (0.46)	3 (2.29)	3 (0.07)	2 (0.03)	3 (0.30)	3 (0.08)	3 (2.99553)
	Junichiro Koizumi	3 (2.7)	3 (1.10)	2 (0.41)	2 (3.42)	2 (0.09)	3 (0.02)	2 (0.32)	2 (0.09)	2 (2.99513)
	Yasuo Fukuda	4 (2.3)	2 (1.17)	4 (0.58)	4 (1.79)	4 (0.02)	4 (0.00)	4 (0.11)	4 (0.03)	4 (2.99860)
Yasuo Fukuda	Takeo Fukuda	1 (9.7)	1 (4.04)	1 (0.16)	1 (11.7)	1 (2.12)	1 (1.20)	1 (2.04)	1 (1.20)	1 (2.99176)
	Tony Blair	2 (4.7)	3 (1.43)	4 (0.52)	3 (1.30)	3 (0.06)	3 (0.01)	4 (0.06)	4 (0.01)	4 (2.99943)
	Nicolas Sarkozy	3 (4.6)	2 (1.75)	2 (0.50)	2 (2.07)	2 (1.03)	2 (0.81)	2 (1.11)	2 (0.82)	2 (2.99518)
	Mamoru Mohri	4 (2.8)	4 (0.73)	2 (0.50)	4 (0.47)	4 (0.01)	4 (0.00)	3 (0.07)	3 (0.02)	3 (2.99886)
Kiichi Miyazawa	Noboru Takeshita	1 (8.4)	1 (3.71)	1 (0.09)	1 (12.1)	1 (1.49)	1 (0.96)	1 (1.85)	1 (1.10))	1 (2.98707)
	George H. W. Bush	2 (4.9)	2 (1.07)	4 (0.58)	3 (0.86)	3 (1.04)	3 (0.81)	3 (1.04)	3 (0.81)	3 (2.99022)
	Robert Rubin	3 (4.0)	4 (0.71)	2 (0.49)	4 (0.46)	4 (0.01)	4 (0.00)	4 (0.02)	4 (0.01)	4 (2.99779)
	Bill Clinton	4 (3.9)	3 (1.05)	2 (0.49)	2 (1.74)	2 (1.06)	2 (0.82)	2 (1.21)	2 (0.86)	2 (2.98931)
Yasuhiro Nakasone	Ronald Reagan	1 (8.5)	1 (1.83)	1 (0.40)	1 (4.98)	1 (1.40)	1 (0.92)	1 (1.53)	1 (0.97)	2 (2.99308)
	Chun Doo-hwan	2 (5.5)	3 (1.40)	3 (0.45)	3 (1.94)	2 (1.21)	2 (0.87)	2 (1.20)	2 (0.85)	3 (2.99408)
	Mikhail Gorbachev	3 (4.0)	2 (1.53)	2 (0.43)	2 (3.22)	4 (0.29)	4 (0.08)	4 (0.28)	4 (0.08)	4 (2.99725)
	Yuri Andropov	4 (3.5)	4 (1.07)	4 (0.51)	4 (0.80)	3 (1.05)	3 (0.82)	3 (1.06)	3 (0.82)	1 (2.99017)
Shigeru Yoshida	Douglas MacArthur	1 (8.3)	1 (2.22)	1 (0.40)	1 (7.23)	1 (1.38)	2 (0.93)	1 (1.58)	1 (0.97)	1 (2.99198)
	John Dulles	2 (5.4)	4 (1.14)	2 (0.47)	3 (1.69)	4 (0.04)	4 (0.01)	4 (0.08)	4 (0.03)	4 (2.99887)
	Harry S. Truman	3 (4.0)	2 (1.37)	4 (0.57)	2 (2.61)	3 (1.08)	3 (0.82)	2 (1.15)	2 (0.84)	3 (2.99311)
	Benito Mussolini	4 (3.3)	3 (1.17)	3 (0.56)	4 (1.59)	2 (1.10)	2 (0.83)	3 (1.08)	3 (0.82)	2 (2.99283)
Taro Aso	Shinzo Abe	1 (8.7)	1 (4.28)	1 (0.15)	1 (25.9)	1 (2.06)	1 (1.18)	1 (3.18)	1 (1.54)	1 (2.98775)
	Condoleezza Rice	2 (5.6)	4 (1.85)	4 (0.50)	4 (2.06)	4 (0.04)	4 (0.01)	4 (1.12)	4 (0.83)	4 (2.99529)
	George W. Bush	3 (4.4)	2 (2.12)	3 (0.48)	2 (4.90)	2 (1.20)	2 (0.86)	2 (1.45)	2 (0.93)	2 (2.99488)
	Kim Jong-il	4 (3.2)	3 (1.99)	2 (0.40)	3 (3.20)	3 (1.11)	3 (0.83)	3 (1.20)	3 (0.85)	3 (2.99512)

In above example, for the source and the destination objects, select famous person known by the participants creating the rankings by their subjects. First select 10 famous Japanese and American politicians as source objects from Japanese Wikipedia, in order to enables the participants to investigate relationships among the persons on Wikipedia and create appropriate rankings. As the destination objects for each source, select four famous persons related to the source. Here only four destinations for each source is selected, because preliminarily observed that participants

sometimes wavered in their judgments for five or more destinations. For each of the 40 obtained pairs of a source and a destination, the strength of the relationship from the source to the destination using this method is computed.

V.   CONCLUSION AND FUTURE WORKS

Thus this method can measure the strength of a relationship between two objects on Wikipedia and rank them. Furthermore, this method does not underestimate objects having high degrees. This paper plans to form a relation tree, a directed tree with a unique node corresponding to the most recent common ancestor of all the entities at the leaves of the tree.

REFERENCES

[1] Y. Koren, S.C. North, and C. Volinsky, "Measuring and Extracting Proximity in Networks," Proc. 12th ACM SIGKDD Int'l Conf.Knowledge Discovery and Data Mining, pp. 245-255, 2006.
[2] M. Ito, K. Nakayama, T. Hara, and S. Nishio, "Association Thesaurus Construction Methods Based on Link Co-OccurrenceAnalysis for Wikipedia," Proc. 17th ACM Conf. Information and Knowledge Management (CIKM), pp. 817-826, 2008.
[3] K. Nakayama, T. Hara, and S. Nishio, "Wikipedia Mining for an Association Web Thesaurus Construction," Proc. Eighth Int'l Conf. Web Information Systems Eng. (WISE), pp. 322-334, 2007.
[4] J. Gracia and E. Mena, "Web-Based Measure of Semantic Relatedness," Proc. Ninth Int'l Conf. Web Information Systems Eng. (WISE), pp. 136-150, 2008.
[5] R.K. Ahuja, T.L. Magnanti, and J.B. Orlin, Network Flows: Theory, Algorithms, and Applications. Prentice Hall, 1993.
[6] K.D. Wayne, "Generalized Maximum Flow Algorithm," PhD dissertation, Cornell Univ., New York, Jan. 1999.
[7] R.L. Cilibrasi and P.M.B. Vita´nyi, "The Google Similarity Distance," IEEE Trans. Knowledge and Data Eng., vol. 19, no. 3, pp. 370-383, Mar. 2007.
[8] G. Kasneci, F.M. Suchanek, G. Ifrim, M. Ramanath, and G. Weikum, "Naga: Searching and Ranking Knowledge," Proc. IEEE 24th Int'l Conf. Data Eng. (ICDE), pp. 953-962, 2008.
[9] F.M. Suchanek, G. Kasneci, and G. Weikum, "Yago: A Core of Semantic Knowledge," Proc. 16th Int'l Conf. World wide Web Conf. (WWW), pp. 697-706, 2007.
[10] "The Erdo¨ s Number Project," http://www.oakland.edu/enp/,2012.
[11] L. Katz, "A New Status Index Derived from Sociometric Analysis," Psychometrika, vol. 18, no. 1, pp. 39-43, 1953.
[12] S. Wasserman and K. Faust, Social Network Analysis: Methods and Application (Structural Analysis in the Social Sciences). Cambridge Univ. Press, 1994.
[13] C. Faloutsos, K.S. Mccurley, and A. Tomkins, "Fast Discovery of Connection Subgraphs," Proc. 10th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 118-127, 2004.
[14] P.G. Doyle and J.L. Snell, Random Walks and Electric Networks, vol. 22. Math. Assoc. Am., 1984.
[15] M. Nakatani, A. Jatowt, and K. Tanaka, "Easiest-First Search: Towards Comprehension-Based Web Search," Proc. 18th ACM Conf. Information and Knowledge Management (CIKM), pp. 2057- 2060, 2009.
[16] L. Finkelstein, E. Gabrilovich, Y. Matias, E. Rivlin, Z. Solan, G.Wolfman, and E. Ruppin, The WordSimilarity-353 Test Collection,2002.
[17] E. Agirre, E. Alfonseca, K. Hall, J. Kravalova, M. Pas¸ca, and A. Soroa, "A Study on Similarity and Relatedness Using Distributional and Wordnet-Based Approaches," Proc. 10th Human Language Technologies: Ann. Conf. North Am. Chapter of the Assoc. Computational Linguistics (NAACL-HLT), pp. 19-27, 2009.
[18] W. Xi, E.A. Fox, W. Fan, B. Zhang, Z. Chen, J. Yan, and D. Zhuang, "Simfusion: Measuring Similarity Using Unified Relationship Matrix," Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 130-137, 2005.
[19] D. Fogaras and B. Ra´cz, "Practical Algorithms and Lower Bounds for Similarity Search in Massive Graphs," IEEE Trans. Knowledge Data Eng., vol. 19, no. 5, pp. 585-598, May 2007.
[20] "Country Ranks 2009," http://www.photius.com/rankings/index.html, 2012.