

LINK PREDICTION USING DISCRETE-TIME QUANTUM WALK

Jing Qian, Lintao Yang, Zetai Yu, Shouyin Liu

Original scientific paper

Link prediction is one of the key issues of complex networks, which attracts much research interest currently. Many link prediction methods have been proposed so far. The classical random walk as an effective tool has been widely used to study the link prediction problems. Quantum walk is the quantum analogue of classical random walk. Numerous research results show that quantum algorithms using quantum walk outperform their classical counterparts in many applications, such as graph matching and searching. But there have been few studies of the link prediction based on quantum walk, especially on discrete-time quantum walk. This paper proposes a new link prediction method based on discrete-time quantum walk. Experiment results show that prediction accuracy of our method is better than the typical methods. The time complexity of our method running on classical computers, compared with the methods based on classical random walk, is slightly higher. But our method can be greatly accelerated by executing on quantum computers.

Keywords: *complex networks; discrete-time quantum walk; computer engineering; link prediction*

Predviđanje veze primjenom kvantnog gibanja u diskretnom vremenu

Izvorni znanstveni članak

Predviđanje veze jedno je od ključnih pitanja složenih mreža koje trenutačno privlači pozor mnogih istraživača. Do sada su predložene mnoge metode predviđanja veze. Klasično slučajno gibanje predstavlja učinkoviti alat koji se uvelike rabi u proučavanju problema predviđanja veze. Kvantno gibanje je kvantni analog klasičnog slučajnog gibanja. Rezultati mnogih istraživanja pokazuju da kvantni algoritmi koji rabe kvantno gibanje nadmašuju svoje klasične kopije u mnogim primjenama, kao što su, na primjer, usklađivanje i istraživanje grafikona. Međutim, malo je istraživanja o predviđanju veze na temelju kvantnog gibanja, a posebice kvantnog gibanja u diskretnom vremenu. U ovom se radu predlaže nova metoda predviđanja veze zasnovana na kvantnom gibanju u diskretnom vremenu. Rezultati eksperimenta pokazuju da je točnost predviđanja našom metodom bolja nego tipičnim metodama. Vremenska složenost naše metode koja se izvodi na klasičnim računalima, u usporedbi s metodama baziranim na klasičnom slučajnom gibanju, malo je bolja. No, naša se metoda može znatno ubrzati izvođenjem na kvantnim računalima.

Ključne riječi: *kvantno gibanje u diskretnom vremenu; predviđanje veze; računarstvo; složene mreže*

1 Introduction

Link prediction is one of the fundamental issues of complex networks. The goal of link prediction is to predict the status of links for unobserved pairs of nodes based on a partially observed network; or to predict the network state at the next time step based on a sequence of fully observed network at various time steps [1]. Link prediction is useful in broad application fields [2, 3], such as recommender systems, protein-protein interaction networks, scientists' collaborations network, the detection of unknown links in terrorism networks, the prediction of friendship formations and the prediction of web hyperlinks. A large number of methods have been proposed to predict links in numerous fields [4]. A good summary is given by Lv and Zhou [1]. The simplest framework of link prediction methods is the similarity-based algorithm. It defines a measure of proximity or similarity between two nodes in the network based on the observation that the more similar nodes have higher probability to be connected. The widely applied methods to measure similarities between nodes are based on random walk on the network. In Average Commute Time (ACT) [5], the similarity between two nodes is defined as the reciprocal of the average number of steps required by a random walker starting from one node to another node. Random Walk with Restart (RWR) [5] is a direct application of PageRank [6]. It considers a random walker starting from one node will iteratively move to a random neighbour with probability c and return to the node with probability $1-c$. SimRank [7] measures how soon two random walkers, respectively starting from one node, are expected to meet at a certain node. Although these

prediction methods provide good accuracy, the time complexities are too high to be practicable for large-scale networks. Liu and Lv [8] puts forward two novel indices, Local Random walk (LRW) and Suspended Random Walk (SRW). It is shown that the prediction accuracy is improved and the computational complexity is reduced, especially for the sparse network with lower clustering Coefficient [8].

Quantum walk, which was first proposed by Aharonov et al. [9], is the quantum analogue of classical random walk. In both, a particle (or walker) in a network moves through the network with time elapsing. In classical random walk, the motion of a particle is diffusive. While in quantum walk, it allows for quantum superposition of the particle states and explores the interference of the terms in these superpositions; and the motion of the particle is more akin to wave propagation [10]. Such differences have motivated a search for quantum algorithms based on quantum walk. Numerous research results show that quantum algorithms outperform their classical counterparts in many applications such as search, element distinctness and combinatorial optimization. For example, Sánchez-Burillo [11] defines a quantum navigation method providing a unique ranking of the elements of a network and solves degeneracy found in the ranking of PageRank. But few of the research works are about the link prediction based on quantum walk.

In this paper, two new node similarity indices based on quantum walk are defined. Like classical random walk, there are two basic types of quantum walk, continuous-time quantum walk (CTQW) [12] and discrete-time quantum walk (DTQW) [9]. Here we consider a particular

version of the discrete-time quantum walk known as the scattering quantum walk (SQW) [13]. Experiment results show that prediction accuracy of our method is better than of the typical methods. The time complexity of our method running on classical computers, compared with the methods based on classical random walk, is slightly higher. But our method can be greatly accelerated by executing on quantum computers.

2 Similarity based on SQW

Let $G = (V(G), E(G))$ be a complex network that is unweighted and undirected, where $V(G)$ is the set of N nodes (or vertices) and $E(G)$ is the set of M undirected edges (or links). Multiple links and self-connections are not allowed. Let $\Gamma(u)$ be the neighbouring nodes of node u and $k(u) = |\Gamma(u)|$ be the total number of neighboring nodes of node u .

SQW is analogous to the propagation of light through an interferometer. In this type of quantum walk, the particle resides on the directed edges and scatters at the nodes at each time step. Each edge $(u, v) \in E(G)$ has two orthogonal states, $|uv\rangle$, corresponding to the particle being on the edge (u, v) and going from node u to node v , and $|vu\rangle$, corresponding to the particle being on the edge (v, u) and going from node v to node u . Set $D(G) = \{uv, vu \mid (u, v) \in E(G)\}$. Let $H \in \mathbb{C}^{2M}$ be Hilbert space spanned by basis states $|uv\rangle$, where $uv \in D(G)$. In general, the state of a SQW, $|\psi\rangle \in H$, is a unit length, complex-linear sum of basis states, which is denoted as

$$|\psi\rangle = \sum_{v \in \Gamma(u)} \alpha_{uv} |uv\rangle \tag{1}$$

where $\alpha_{uv} \in \mathbb{C}$ is the quantum amplitude of the state $|uv\rangle$. The probability that the walk is in the state $|uv\rangle$ is given by $p(|uv\rangle) = \alpha_{uv} \alpha_{uv}^*$, where α_{uv}^* is the complex conjugate of α_{uv} .

Unlike classic random walk, the evolution of SQW is governed by unitary matrix

$$|\psi(n+1)\rangle = U|\psi(n)\rangle \tag{2}$$

where U is the quantum amplitude matrix which determines the one-step transition probabilities for transitions between different states. U is a unitary matrix, i.e. $U^{-1} = (U')^*$. The sum of the squares of the amplitudes for all the transitions from a particular state must be unity. Suppose we are in the state $|\psi\rangle = |uv\rangle$. If the particle is transmitted, it will be in the adjacent states $|vx\rangle$ ($x \in \Gamma(v) \setminus u$) and if reflected in the state $|vu\rangle$. Let the transmitted amplitude be t , and the reflection amplitude be r . Then the transition will be of the form

$$|uv\rangle \rightarrow r|vu\rangle + t \sum_{x \in \Gamma(v) \setminus u} |vx\rangle \tag{3}$$

Here we use Grover diffusion matrices [14], which assigns quantum amplitudes of $r = \frac{2}{k(v)} - 1$ and $t = \frac{2}{k(v)}$.

The Grover diffusion matrix $U = U(G) = (U_{uv,wx})_{uv,wx \in D(G)}$ of G is defined by:

$$U_{uv,wx} = \begin{cases} \frac{2}{k(x)} - 1 & u = x, v = w \\ \frac{2}{k(x)} & u = x, v \neq w \\ 0 & \text{else} \end{cases} \tag{4}$$

Obviously, a particle can make a transition from one directed edge uv to another directed edge wx if and only if $v = w$ and $v \neq x$.

Given a quantum walker starting from node u , we consider that there is an equal prior probability to find the quantum walker on any directed edge uv ($v \in \Gamma(u)$) at the start of evolution (at $n = 0$). This leads to an initial state of the form

$$|\psi(0)\rangle = \sum_{v \in \Gamma(u)} \alpha_{uv}(0) |uv\rangle \tag{5}$$

where $\alpha_{uv}(0) = \frac{1}{\sqrt{|\Gamma(u)|}}$. After n step, the evolution of the SQW is given by

$$|\psi(n)\rangle = U^n |\psi(0)\rangle \tag{6}$$

Let $p_{ux}(n)$ be the probability of visiting a node x after n step of the quantum walk on G starting at u . The probability $p_{ux}(n)$ is defined as

$$p_{ux}(n) = \sum_{w \in \Gamma(x)} |\langle xw | \psi(n) \rangle|^2 \tag{7}$$

In general p_{ux} does not converge either. The reason being that U , as a unitary transformation, prevents the walk from reaching a steady state.

We modify $p_{ux}(n)$ to $p'_{ux}(n) = q_u p_{ux}(n)$ by considering the initial configuration function q_u . Here q_u is defined as $\frac{|\Gamma(u)|}{2M}$, which is proportional to the degree of node u . Herein, we propose a new link prediction index, local SQW index $s_{ux}^{LSQW}(n)$, mixing both $p_{ux}(n)$ and $p_{xu}(n)$ to characterize more efficiently the similarity between node u and x :

$$s_{ux}^{LSQW}(n) = \frac{p'_{ux}(n) + p'_{xu}(n)}{1 + p'_{ux}(n)p'_{xu}(n)} = \frac{q_u p_{ux}(n) + q_x p_{xu}(n)}{1 + q_u q_x p_{ux}(n)p_{xu}(n)} \tag{8}$$

where $q_u = \frac{|\Gamma(u)|}{2M}$ and $q_x = \frac{|\Gamma(x)|}{2M}$. It is obvious that $s_{ux} = s_{xu}$.

One difficulty with LSQW index, which is the same as all random walk based indices, is its sensitive dependence on parts of the network far from u and x , even when u and x are connected by very short paths [1]. Away of counteracting this dependence is to continuously release the walkers at the starting point independently; in this way, the target node and the nodes nearby lead to higher prediction accuracy. Similarly to [8], the corresponding index, superposed SQW index $s_{ux}^{SSQW}(n)$, is

$$s_{ux}^{SSQW}(n) = \sum_{l=1}^n s_{ux}^{LSQW}(l) \quad (9)$$

3 Experiment

In this section, we first describe our experiments on real data sets, and then present the performance of several baselines and our link prediction methods.

3.1 Data sets

We import six typical real-world complex networks collected from different fields: (i) Lesmis: Lesmis is the coappearance network of characters in the novel Les Miserables; (ii) NS (netscience): A network of coauthorships between scientists who are themselves publishing on the topic of network science; (iii) C.elegans: The neural network of the nematode worm C.elegans, in which an edge joins two neurons if they are connected by either a synapse or a gap junction; (iv) jazz: a collaboration network of jazz musicians; (v) USAir: The network of the US air transportation system. The weight of a link is the frequency of flights between two airports; (vi) email: the e-mail network of University Rovira i Virgili (URV) in Tarragona, Spain. The detailed statistics descriptions of these data sets are listed in Tab. 1, where N and M are the total numbers of nodes and links, respectively; $\langle k \rangle$ is the average degree of the network; $\langle d \rangle$ is the average shortest distance between node pairs; C is the clustering coefficient.

Table 1 The basic statistics features of the six data sets

Datasets	N	M	$\langle k \rangle$	$\langle d \rangle$	C
lesmis	77	254	6.6	2.64	0.5731
NS	379	914	4.82	6.04	0.7412
C.elegans	297	2148	14.46	2.46	0.2924
jazz	198	2742	27.7	2.24	0.6175
USAir	332	2126	12.8	2.74	0.6252
email	1133	5451	9.62	3.6	0.22

3.2 Evaluation metrics

To quantify the prediction accuracy of each method, we adopt the approach used in [15]. In $G = (V, E)$, the link set E is randomly divided into two nonoverlapping parts: the training set E^{train} and the test set E^{test} . Here, E^{train} is assumed to be known information and E^{test} is the

testing set. It is obvious that $E^{train} \cup E^{test} = E$ and $E^{train} \cap E^{test} = \emptyset$. The prediction results are the set of missing links or link to occur in the future based on E^{train} . The prediction accuracy is obtained by comparing the prediction results with E^{test} .

We use AUC [15] to quantify the accuracy of prediction measures. The notion of AUC is the probability that a randomly chosen missing link (a link in E^{test}) is given a higher score than a randomly chosen nonexistent link (a link not in E). In k times independent comparisons, if there are k' times the missing link having a higher score and k'' times they are of the same score. The calculation of AUC is $AUC = \frac{k' + 0.5k''}{k}$. When all the scores are generated from an independent and identical distribution, the AUC should be about 0.5. The degree to which the AUC exceeds 0.5 shows how much better the algorithm performs than pure chance.

3.3 Baseline predictors

For the sake of comparison, the link prediction results on five baseline predictors are presented. Two indices based on classical random walks, Local Random Walk and Superpose Random Walk index, are considered. In addition, in [16] the Common Neighbours index, the Adamic-Adar index [17] and Resource Allocation index [18] had better performance in most cases. Here the three methods are also considered as the basic predictors. A brief introduction of these indices is given as follows.

(1) Common Neighbours (CN):

$$S_{xy}^{CN}(x, y) = |\Gamma(x) \cap \Gamma(y)| \quad (10)$$

where $\Gamma(x)$ is the set of neighbours of x in G . The CN predictor is according to the observation that two nodes are more similar with more common neighbours.

(2) Adamic-Adar (AA):

$$S_{xy}^{AA}(x, y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log(k(z))} \quad (11)$$

where $k(z)$ is the degree of the node z . The AA predictor is according to the observation that the common neighbor which has a small degree contributes more similarity than the one that has a large degree. This index assigns rarer connected node more weights.

(3) Resource Allocation Index (RA):

$$S_{xy}^{RA}(x, y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{k(z)} \quad (12)$$

This index is motivated by the resource allocation dynamics on complex networks.

(4) Local Random Walk (LRW):

This metric is based on the classical random walk. The transition probability matrix is P , with $P_{xy} = 1/k(x)$. The P_{xy} presents the probability that a random walker staying at node x will walk to y in the next step. Suppose a

random walk starting at x , the probability that it locates at y after n steps is $\pi_{xy}(n)$. The similarity based on Local Random Walk is

$$S_{xy}^{LRW}(n) = \frac{k_x}{2M} \pi_{xy}(n) + \frac{k_y}{2M} \pi_{yx}(n) \quad (13)$$

(5) Superposed Random Walk (SRW):

In order to counteract the high clustering or locality of networks which may lead to nodes' sensitive dependence on parts of the network far away from target nodes, Superposed Random Walk Index superposes the contribution of independently moved walkers. The SRW similarity index is defined as

$$S_{xy}^{SRW}(n) = \sum_{l=1}^n S_{xy}^{LRW}(l) \quad (14)$$

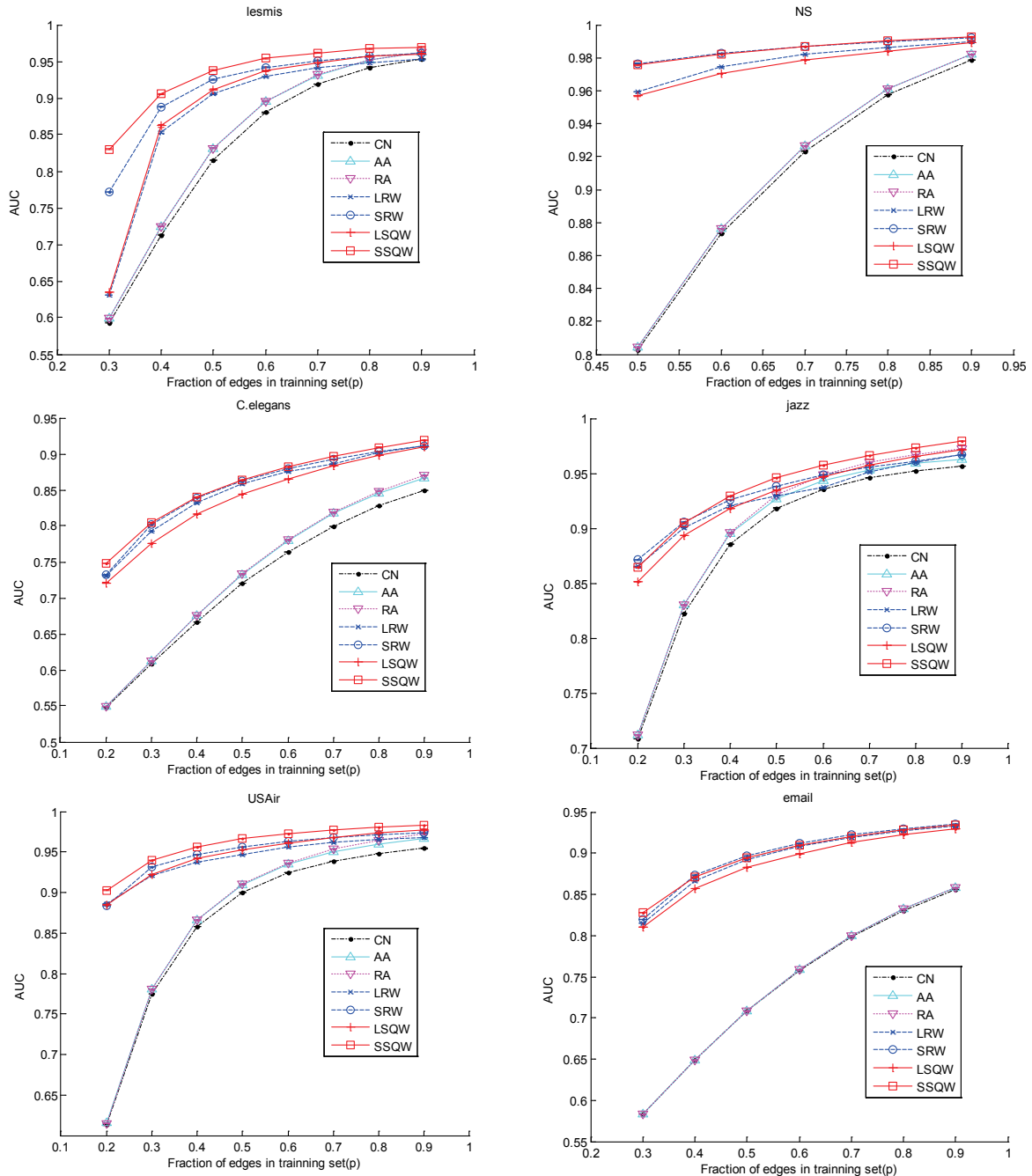


Figure 1 Dependence of AUC on the size of training set in USAir. Each number is obtained by averaging over 1000 implementations with independently random divisions of the training set and probe set. The training set represents a connected network. The accuracy of random walk and quantum walk is generated by the corresponding optimal step of each of the indices.

3.4 Methods comparison

In our first experiment, we perform the seven methods on six real-world networks with various fractions of edges in training set, p . The results are shown in Fig. 1.

It can be seen that while the p increases, all these similarity-based methods have better prediction performance. In network with low density (for example $p \leq 0.5$) the methods based on random walk and quantum walk perform better than the other three typical methods.

And SSQW seem to be the best method on most networks while p grows to a certain value, except for on NS and email. But the optimal results of NS and email are very close to the corresponding AUC curves of SSQW. Moreover, it is also observed that similar results as [8], LSQW and SSQW are not sensitive to the density of the network.

As SRW (or LRW), our methods also have an important parameter n (the number of steps). We now discuss how to set this parameter and how the parameter setting affects the link prediction results. Fig. 2 shows the correlation between the number of steps and the value of AUC. For all data sets, when the number of step is small, e.g. $n \leq 4$, the link prediction performance of the link prediction performance of SSQW (or LSQW) and SRW (or LRW) is close. As n increases, SSQW (or LSQW)

performs better than SRW (or LRW). The two exceptions are the prediction results of LSQW and LRW on NS and the prediction results of SSQW and SRW on email, but the two curves in each pair of curves are rather close to each other. Meanwhile, we observe that the link prediction accuracy of SSQW (or LSQW) decreases more slowly with the increase of steps than that of SRW (or LRW). Especially, the AUC value of SSQW becomes stable after $n \geq 6$. The results indicate that SSQW is not sensitive to the number of the steps, while SRW is sensitive to the number of the steps. It is because of the time-reversal invariance of a quantum walk [13]. Although [8] mentions the positive correlation between the optimal step and the average shortest distance in SRW, it should still be careful to select the value of the step. In SSQW, we set the value of n to be 7.

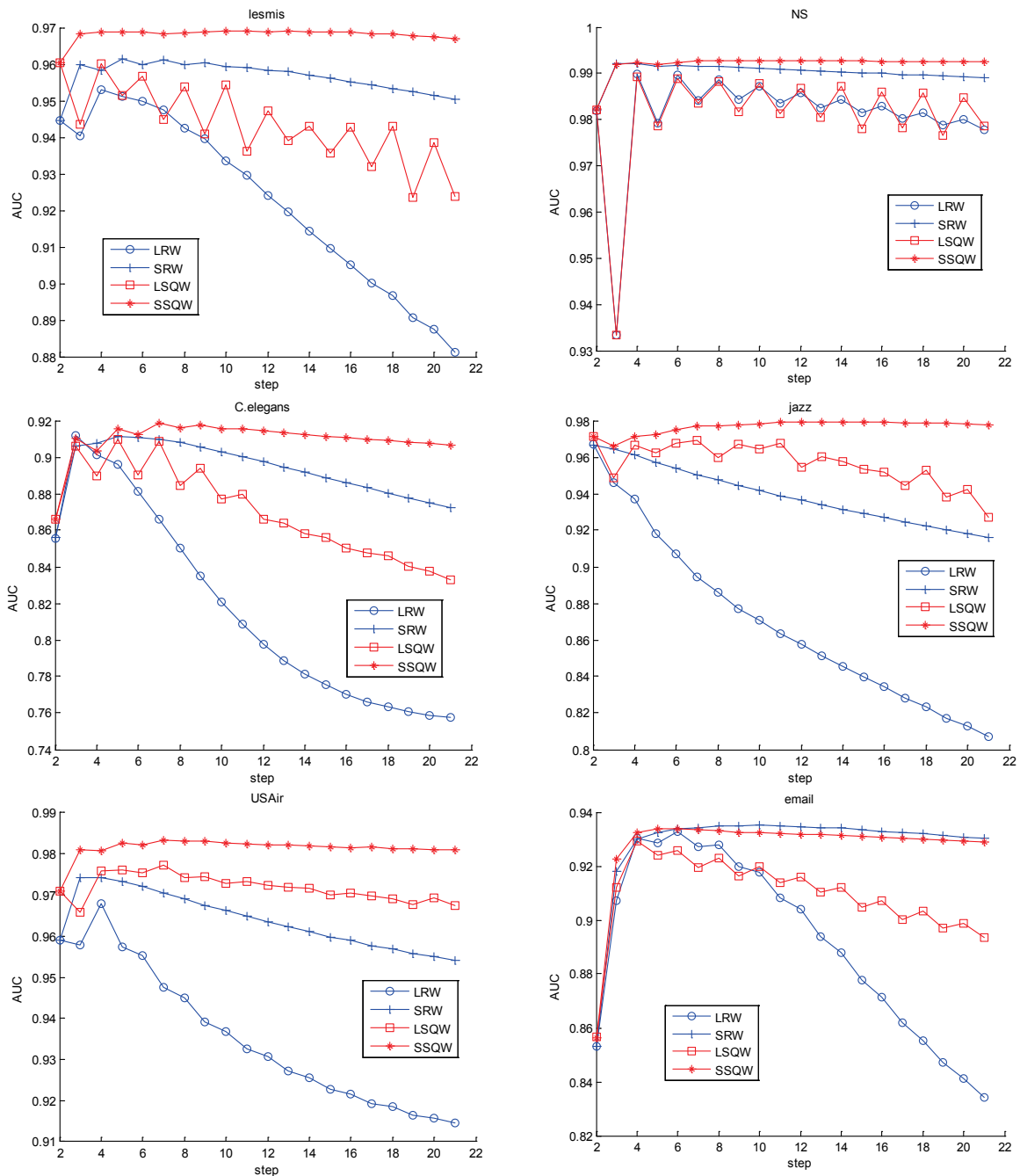


Figure 2 The effect of the number of steps. For each network, the training set contains 90% of the known links. Each number is obtained by averaging over 1000 implementations with independently random divisions of training set and probe set

Although our methods use the aspects of quantum mechanics such as interference and superposition, they can be readily implemented on classical computer. Since the number of states for the SQW is $2M$, the complexity of our methods is $O(M\langle k \rangle^n)$. As we known, the time complexity of n -step SRW (or LRW) is approximately $O(N\langle k \rangle^n)$. In terms of raw computations required, our method is slightly more expensive than SRW (or LRW) on classical computer. However, the simulation of the discrete-time quantum walk achieves quadratic speed up via Grover's technique [14] and this means, that the simulation of the quantum walk could be realized with complexity $O(\sqrt{M\langle k \rangle^n})$ on quantum computer. It is obvious that the time complexity of SSQW (or LSQW) on quantum computer is much less than that of SRW (or LRW) on classical computer.

4 Conclusions

In this Letter, a quantum perspective was applied to the solution of link prediction problem, trying to investigate whether quantum walk may improve the prediction accuracy, compared with classical random walk. We first proposed two similarity indices based on SQW: the LSQW and the SSQW index. Our methods were compared with five well-known methods on six real networks. The results show that SSQW can give overall better predictions than the other methods on AUC scores. Otherwise, the time complexity of SSQW (or SLQW) on classical computers is a little higher than that of SRW (or LQW). But SSQW (or SLQW) gives rise to a quadratic speedup on quantum computer.

Our work can be further improved in several aspects. First, we will study in depth how our methods perform on networks with different structural properties. Second, CTQW's behaviour can consequently be markedly different to that of the DTQW. We plan to study the link prediction on continuous-time quantum walk in future work.

Acknowledgements

The authors would like to acknowledge the reviewers for their useful comments and advices. This work was supported by the National Natural Science Foundation of Hubei province under Grant No.2013CFB210.

5 References

- [1] Lv, L.; Zhou, T. Link prediction in complex networks: A survey. // *Physica A*. 390, 6(2011), pp. 1150-1170. <https://doi.org/10.1016/j.physa.2010.11.027>
- [2] Jiang, Y.; Hua, Z.; Li, Z. A new compound faults detection method for rolling bearings based on empirical wavelet transform and chaotic oscillator. // *Chaos, Solitons and Fractals* (2015). <https://doi.org/10.1016/j.chaos.2015.09.007>
- [3] Li, Z.; Peng, Z. A new nonlinear blind source separation method with chaos indicators for decoupling diagnosis of hybrid failures: A marine propulsion gearbox case with alarge speed variation. // *Chaos, Solitons and Fractals* (2015). <https://doi.org/10.1016/j.chaos.2015.09.023>
- [4] Li, Z.; Jiang, Y.; Wang, X.; Peng, Z. Multi-mode separation and nonlinear feature extraction of hybrid gear failures in coal cutters using adaptive nonstationary vibration analysis. // *Nonlinear Dynamics* (2015) <https://doi.org/10.1007/s11071-015-2505-3>
- [5] Fouss, F.; Pirotte, A.; Renders, J.; Saerens, M. Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation. // *IEEE Trans. Knowl. Data. Eng.* 19(2007), pp. 355-369. <https://doi.org/10.1109/TKDE.2007.46>
- [6] Brin, S.; Page, L. The Anatomy of a Large-Scale Hypertextual Web Search Engine. // *Computer Network and ISDN Systems*. 30 (1998), pp. 107-117. [https://doi.org/10.1016/S0169-7552\(98\)00110-X](https://doi.org/10.1016/S0169-7552(98)00110-X)
- [7] Jeh, G.; Widom, J. SimRank: A measure of structural-context similarity. // In *Proceedings of 8th ACM SIGKDD International Conference on knowledge Discovery and Data Mining*, Edmonton, Alberta, Canada, (2002), pp. 271-279. <https://doi.org/10.1145/775047.775126>
- [8] Liu, W.; Lv, L. Link prediction based on local random walk. // *EPL (Europhysics Letters)*. 89, 5 (2010), pp. 58007. <https://doi.org/10.1209/0295-5075/89/58007>
- [9] Aharonov, Y.; Davidovich, L.; Zagury, N. Quantum random walks. // *Physics Review A*. 48, 2(1993), pp. 1687-1690. <https://doi.org/10.1103/PhysRevA.48.1687>
- [10] Kempe, J. Quantum random walks-an introductory overview. // *Quantum Physics*. 44(2003), pp. 307-327. <https://doi.org/10.1080/00107151031000110776>
- [11] Sánchez-Burillo, E.; Duch, J.; Gómez-Gardeñes, J.; Zueco, D. Quantum navigation and ranking in complex networks. // *Scientific Reports*. 2, 605(2012), pp. 1-8. <https://doi.org/10.1038/srep00605>
- [12] Mülken, O.; Blumen, A. Continuous-time quantum walks: Models for coherent transport on complex networks. // *Physics Reports*. 502, 2(2011), pp. 37-87. <https://doi.org/10.1016/j.physrep.2011.01.002>
- [13] Feldman, E.; Hillery, M. Modifying quantum walks: a scattering theory approach. // *J. Phys. A: Math. Theor.* 40, (2007), pp. 11343-11359. <https://doi.org/10.1088/1751-8113/40/37/011>
- [14] Grover, L. A fast quantum mechanical algorithm for database search. // In *Proceedings of 28th Annual ACM Symposium on the Theory of Computation*, Philadelphia, Pennsylvania. 1996, (1996), pp. 212-219. <https://doi.org/10.1145/237814.237866>
- [15] Hanely, J.; McNeil, B. The meaning and use of the area under a receiver operating characteristic (ROC) curve. // *Radiology*. 143, 1(1982), pp. 29-36. <https://doi.org/10.1148/radiology.143.1.7063747>
- [16] Liben-Nowell, D.; Kleinberg, J. The link-prediction problem for social networks. // *Journal of the American Society for Information Science and Technology*. 58, 7(2007), pp. 1019-1031. <https://doi.org/10.1002/asi.20591>
- [17] Adamic, L.; Adar, E. Friends and neighbors on the web. // *Social Networks*. 3(2003), pp. 211-230. [https://doi.org/10.1016/S0378-8733\(03\)00009-1](https://doi.org/10.1016/S0378-8733(03)00009-1)
- [18] Zhou, T.; Lv, L.; Zhang, Y. Predicting missing links via local information. // *The European Physical Journal B*. 71, 4 (2009), pp. 623-630. <https://doi.org/10.1140/epjb/e2009-00335-8>

Authors' addresses

Jing Qian, PhD
Lintao Yang, PhD
Zetai Yu, PhD
Shouyin Liu, PhD
 College of Physical Science and Technology,
 Central China Normal University,
 No. 152, Luoyu Road, Wuhan 430079, China
 E-mail: jingqian_ccnu@163.com