

Social Transfer Cross Domain Transfer Learning From Social Streams for Media Applications

Harikrishna Chaganti¹, G. Ramachandrarao²

¹M.Tech Student, Department of Computer Science & Engineering
Chalapathi Institute of Engineering & Technology, Guntur, Andhra Pradesh, India.

²Assistant Professor, Department of Computer Science & Engineering
Chalapathi Institute of Engineering & Technology, Guntur, Andhra Pradesh, India.

Email: harikrishnachaganti@gmail.com, grrsoft30@gmail.com.

Abstract—Recommender systems can suffer from data scarcity and cold start issues. However, social networks, which enable users to build relationships and create different types of items, present an unprecedented opportunity to alleviate these issues. In this paper, we represent a social network as a star-structured hybrid graph centered on a social domain, which connects with other item domains. With this innovative representation, useful knowledge from an auxiliary domain can be transferred through the social domain to a target domain. Various factors of item transferability, including popularity and behavioral consistency, are determined. We propose a novel Hybrid Random Walk (HRW) method, which incorporates such factors, to select transferable items in auxiliary domains, bridge cross-domain knowledge with the social domain, and accurately predict user-item links in a target domain. Extensive experiments on a real social dataset demonstrate that HRW significantly outperforms existing approaches.

Keywords: Cross-domain media retrieval, recommendation, transfers learning, social media.

1. Introduction:

A social networking service is a platform on which users can create and adopt different types of items such as web posts (e.g., articles and tweets), user labels, images, and videos. The huge volume of items generates a problem of information overload. Traditional web post recommendation approaches suffer from data scarcity (i.e., limited interaction between users and web posts) and the issue of cold start (i.e., giving recommendations to new users who have not yet created any web posts). The social connections and multiple item domains found in social networks provide an unprecedented opportunity to alleviate these issues in real applications.

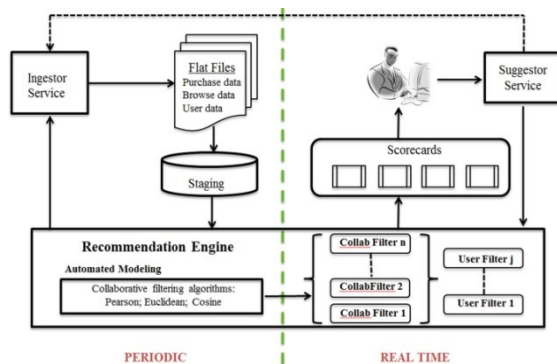
One common type of approach to recommendations, known as collaborative altering

(CF) techniques, characterizes users' latent features independently with user-item interactions in a single item domain [1]. Similarly, the type of approach provided in [2] does not consider the question of multiple domains. However, users' characteristics relate both to social connections and to different user-item interactions. For example, users read web posts created by their community and may adopt similar user labels to their friends. Therefore, an effective social recommendation approach should acknowledge (1) social tie strength (hence forth, tie strength) between users and (2) different user item interactions. The problem of how to incorporate a social domain and auxiliary item domains (e.g., user labels and images) into a unified framework remains open. Frameworks exist that connect directly-related item domains, such as a music album and tags on that album [3], or web pages and queries to them [4]. However, these cannot be applied to indirectly-related item domains in social networks, such as tweets and user labels. The multiple item domains reflect users' intrinsic preferences and tend to be tightly connected among a massive number of users. In this paper, we reconsider the representation of social networks and propose a star-structured graph, where the social domain is at the center and is connected to the surrounding item domains, as shown in Fig. 1.

The value of the cross-domain link1 weight represents how often a given user adopts a given item, while the value of the within-domain link2 weight in the social domain represents the tie strength between users. Tie strength can refer to homophile [5], circle-based influence [6] [7] [8] or social trust [9], [10]. Users are more likely to have stronger ties if they share similar characteristics. Cross-domain links reflect users' characteristics in different ways. For example, a cross-domain link from a user to a web post about iPhones shows his/her short-term interest in iPhones, and a cross-domain link from him/her to a label "iPhone Fan" implies his/her long-term interest in iPhones. A basic assumption is that the more auxiliary knowledge we have, the more we know about the

users, thereby enabling more accurate estimates of tie strength. When a user and his/her friend have many common user labels, we assume greater tie strength and expect them to be more similar in terms of their web post adoption behaviors. Even if the web post domain is extremely sparse, we may still produce effective recommendations by transferring auxiliary knowledge from other item domains through the social domain.

Thus, knowledge transfer procedures among multiple item domains in social networks should focus on updating tie strength in the social domain, but this is complicated by challenges associated with jointly modeling multiple relational domains, discovering transferable knowledge, and improving recommendations in the target domain.



The following characteristics of the domains considered in this paper are challenging to deal with when developing approaches to recommendation are:

1. The domains are relational. Social network data provide social connections between users, semantic similarity between two items of the same type, and item adoptions by users. The issue of how to represent the user-user links, item-item links, and user-item links poses a challenge to method capability.
2. The domains are heterogeneous. Heterogeneity is a challenging issue in social recommendation. Within domain links can be directed (“following” links in the social domain) or undirected (semantic similarity links in the item domains). Cross-domain links can be signed (indicating a positive or negative connotation, such as web-post adoptions and rejections) or unsigned (user-label adoptions). The issue of how to transfer knowledge across heterogeneous domains poses a challenge to method comprehensibility.
3. The domains are variously sparse. This data scarcity is essentially caused by the large amounts of users and items as well as the time and attention scarcity of these users. It is challenging to try to use relatively dense

auxiliary information to help predict sparse links in the target domain.

4. Items in the domains have varying transferability. Traditional literature often assumes that the most popular items have better transferability. However, later in this work, we will show that this assumption is incorrect. Therefore, transferable knowledge selection approaches for enhancing performance constitutes a literature gap.

To address the above challenges, we propose an innovative hybrid random walk (HRW) method for transferring knowledge from auxiliary item domains according to a star structured configuration to improve social recommendations in a target domain. HRW estimates weights for (1) links between user nodes within the social domain, and (2) links between user nodes in the social domain and item nodes in the item domain. The weights respectively represent (1) tie strength between users and (2) the probability of a user adopting or rejecting an item. Our proposed method integrates knowledge from multiple relational domains and alleviates scarcity and cold-start issues. The key contributions are:

1. We discover counterintuitive transferability distribution in auxiliary item domains. Besides popularity, we find more meaningful factors, i.e., behavioural consistency with web post adoptions and social connections. These factors have been incorporated into our method.
2. We propose a novel method to transfer knowledge across multiple relational domains on social networks, incorporating heterogeneous graphs with different types of links. This method can be naturally applied to graph-based applications such as social networks, information networks, and biological networks.
3. Extensive experimentation on a large real social dataset demonstrates that HRW produces significantly superior recommendations for web posts on social networks. In terms of providing recommendations to cold-start users, only 30 percent of historical data from the web-post domain is necessary to achieve a comparable performance to that of an approach that makes use of user-label data.

The remainder of this paper is organized as follows. Section 2 discusses related works. Section 3 provides some background and preliminary concepts regarding transfer ability. Section 4 describes the methodology of the HRW approach with Section 5 providing experimental results. Section 6 concludes.

2. RELATED WORK

In this section, we survey related works and note the literature gap that our research addresses.

2.1 Cross-Domain Collaborative Filtering

Collaborative filtering techniques are a common approach to recommendation, and have been applied in real recommender systems [11], [12], [13]. Based on probabilistic matrix factorizations [1], approaches have been proposed to improve recommendation by jointly factorizing a trust network and a user-item matrix [2], [14]. One particular contextual model learns both individual preference and interpersonal influence to estimate the probability of item adoption [15]. However, when dealing with information overload, CF does not consider the interplay of users and multiple types of items, so it often suffers from data scarcity and cold start issues.

Recommender systems benefit from new information that goes beyond the user-item matrix. Berkovsky et al. Deployed several mediation approaches for importing and aggregating user rating vectors from different domains. Gao et al. [21] conducted recommendation via a cluster-level latent factor model. A joint model of tensor factorization was proposed by Chen et al. to simultaneously recommend users, movies, and tags [22]. However, social applications are different from movie recommenders. Social relation drives information diffusion and adoption [23], [24]. Only with the consideration of tie strength can recommender systems better understand users' behavioral intentions. Social Matchbox [25] proposed a latent factor matrix factorization model that treats users' side information (user profiles) and social information as feature vectors to determine user

Similarity. Facebook uses cross-domain data (user profiles and new feeds) for their recommenders [26]. Sedhain et al. noted that side information is very important [27]. Auxiliary item domains are more complicated than side information. Rich auxiliary user-item interactions, including editing user labels, sharing videos, and joining groups, should be incorporated into a more relational, random walk-based model than factorization-based methods. Fortunately, the advantage of random walk models in utilizing auxiliary information has been proved by empirical research [28].

2.2 Transfer Learning for Recommendation

Adomavicius and Tuzhilin [29] reviewed CF based, content-based, and hybrid recommendation methods. Their work predicts that auxiliary

information will play an important role in the future of recommender systems. Transfer learning provides the key idea of using knowledge from auxiliary domains [30] [31] [32] [33] [34][35] and has been used in various ways. Transferring collaborative knowledge from Movie Lens can reduce the sparsity problem in recommending movies in Netflix [36]. Book ratings and movie ratings can also be used collaboratively: transferring book ratings can improve movie rating prediction [37]. A recent work by Jing et al. provides a probabilistic collective factorization model to handle sparse data in different settings of knowledge transfer [38]. Transfer learning methods often utilize users' consistent individual preference to bridge two domains by the set of user nodes. However, in social networks, tie strength between users is the key factor utilized to bridge two item domains. We reconsider the representation of social networks, using a hybrid star-structured graph to incorporate within-domain and cross-domain links. Using the social domain as the bridge between the item domains is unique among existing works.

2.3 Random Walk Algorithms and Models

The random walk concept has been widely applied in recommender systems. Random walk based approaches effectively incorporate auxiliary information [28]. Tong et al. [39] proposed a computationally efficient random walk algorithm. Item Rank [40], a random walk-based scoring algorithm, was used to rank products according to expected user preferences. TrustWalker [10] defined and measured the confidence of a recommendation with a random walk model to

RELATED WORKS

combine the trust-based and CF approaches. Chen et al. [28] proposed a random walk algorithm to handle both positive and negative comments with the guarantee of convergence. In social networks, social relations and multiple item domains naturally form a star-structured high order heterogeneous graph [41], [42]. In this paper, we develop a random walk-based algorithm on such complex graphs to transfer knowledge from rich, auxiliary domains to a target domain. The biggest difference with respect to previous works is that we use social ties as the fundamental bridge to connect item domains in social networks.

Algorithm 1 *SocialTransfer*: Transfer Learning from Social Stream

Input: A target classification task which includes the target training data χ_{train} , the source auxiliary data χ_{aux} , and the target test data χ_{test} .

Output: Classification result on χ_{test}

- 1: Construct the initial transfer graph $G(V, E)$ based on the social transfer clustering task (c.f. Section 3.2).
- 2: Calculate transfer Laplacian matrix: L_{input} from G by Eq. (3).
- 3: **for** each chunk of tweets entering the system **do**
- 4: Calculate the regularization vector δ_1 using the input supervision of social topics $A^{(n)}$ by Eq. (4).
- 5: Perform semi-supervised topic bias update on transfer Laplacian: $L_{topic_bias} = L_{input} + \gamma \cdot \delta_1 \delta_1^T$ by Eq. (5).
- 6: Use Power Iteration to calculate the first q eigenvectors of L_{topic_bias} : E_1, E_2, \dots, E_q which satisfy the generalized eigenproblem: $L_{topic_bias} E = \lambda U E$. The resulting eigenvectors will be used as initial eigenvectors for the next updated Laplacian matrix.
- 7: **end for**
- 8: Construct matrix H with E_1, E_2, \dots, E_q as columns.
- 9: **for** each $\chi_{tr}^m \in \chi_{train}$ **do**
- 10: Let u_{tr}^m be the corresponding row in H w.r.t χ_{tr}^m .
- 11: **end for**
- 12: Use a traditional classification algorithm (we use SVM) to train the classifier $f^l(\chi_{test})$ based on $U_{tr} = \{u_{tr}^m\}_{m=1}^M$ instead of the original training set $\chi_{train} = \{x_{tr}^m\}_{m=1}^M$ and then classify $\chi_{test} = \{x_{ts}^n\}_{n=1}^N$ in the eigen feature space.

3. PRELIMINARIES ON TRANSFERABILITY

In this section, we introduce our social dataset of multiple item domains and demonstrate the existence of transferability from auxiliary domains to a target domain.

3.1 Dataset and Distributions

The dataset for this research was crawled in January 2011 from Tencent Weibo (t.qq.com). We crawled data from users who own at least one user label. While the website allows users to have, at most, 10 user labels, the average number of user labels per user was 5.3. The average number of web posts per user was 12.8. We did not filter any social relationships. The average number of friends per user was 14.2.

Table 1 summarizes the data statistics. We used a 5-minute time window to derive negative links. That is, if a user had two adopting behaviors (sharing the web posts) in 5 minutes, we assumed that the user ignored the rest of posts that he/she received in the time window. Thus, besides the two positive user-post links, we noted several negative links. The data indicates that although both web-post and user-label domains are sparse, the latter is denser.

Fig. 2 shows (1) distributions of user and post frequency, (2) distributions of user and label frequency, and (3) distributions of follower and follow frequency. We note that the data has smooth distributions, which look like power law relationships in log-log scale. Our dataset has no spiky outliers.

Topical Words	Assigned Topic	YouTube Category
dance, adventure, photography, visit	events	Travel & Events
anime, hero, online, celebrity, diva	films	Films & Animation
iphone, games, showcase	electronics	Sci. & Tech
war, economy, army, revolution, blog, egypt	politics	News & Politics
trailer, show, live, watch	entertainment	Entertainment
wow, rap, jam, gaga	music	Travel & Music

4. HYBRID RANDOM WALK ALGORITHM

In this section we introduce our random walk based method on social recommendation. Owing to data scarcity in the target domain, traditional bipartite random walk (BRW) algorithms cannot accurately derive user tie strength to predict user behaviors in the target domain [43], [44]. Fortunately, we have auxiliary domains in which user ties are formed for The same reason as in the target domain: homophile, trust, and influence. The key idea is to utilize rich knowledge from the auxiliary domains to better describe user tie strength and then more precisely predict user behaviors. Thus, we derive HRW algorithms on star-structured graphs.

5. CONCLUSION

In this paper, we addressed the problems of data sparsity and cold start in social recommendation. We reconsidered the problem from the transfer learning perspective and alleviated the data sparsity problem in a target domain by transferring knowledge from other auxiliary social relational domains. By considering the special structures of multiple relational domains in social networks, we proposed an innovative HRW method on a star-structured graph, which is a general method to incorporate complex and heterogeneous link structures.

We conducted extensive experiments on a large real-world social network dataset and showed that the proposed method greatly boosts the social recommendation performance. In particular, we gained improvement in web post recommendation by transferring knowledge from the user-label domain for the user tie strength updating process compared with the recommendation methods, which only use information from the web-post domain.

The proposed method and insightful experiments indicate a promising and general way to solve the data sparsity problem.

6. REFERENCES

1. Y. Koren, R. Bell and C. Volinsky, "Matrix factorization techniques for recommender systems", *Computer*, vol. 42, no. 8, pp. 30-37, 2009
2. H. Ma, D. Zhou, C. Liu, M. R. Lyu and I. King, "Recommender system with social regularization", *Proc. 4th ACM Int. Conf. Web Search Data Mining*, pp. 287-296, 2011
3. I. Konstas, V. Stathopoulos and J. M. Jose, "On social networks and collaborative recommendation", *Proc. 32nd Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, pp. 195-202, 2009
4. J. Wang, H. Zeng, Z. Chen, H. Lu, L. Tao and W. Ma, "ReCoM: Reinforcement clustering of multi-type interrelated data objects", *Proc. Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, pp. 274-281, 2003
5. J. Brown, A. J. Broderick and N. Lee, "Word of mouths communication within online communities: Conceptualizing the online social network", *J. Interactive Marketing*, vol. 21, pp. 2-20, 2007
6. J. Leskovec, A. Singh and J. Kleinberg, "Patterns of influence in a recommendation network", *Proc. 10th Pacific-Asia Conf. Knowl. Discovery Data Mining*, vol. 3918, pp. 380-389, 2006
7. X. Yang, H. Steck and Y. Liu, "Circle-based recommendation in online social networks", *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, pp. 1267-1275, 2012 []
8. X. Qian, H. Feng, G. Zhao and T. Mei, "Personalized recommendation combining user interest and social circle", *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 7, pp. 1763-1777, 2014
9. P. Massa and P. Avesani, "Trust-aware recommender systems", *Proc. ACM Conf. Recommender Syst.*, pp. 17-24, 2007 []
10. M. Jamali and M. Ester, "TrustWalker: A random walk model for combining trust-based and item-based recommendation", *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, pp. 397-406, 2009
11. N. N. Liu, M. Zhao and Q. Yang, "Probabilistic latent preference analysis for collaborative filtering", *Proc. 18th ACM Conf. Inf. Knowl. Manage.*, pp. 759-766, 2009
12. M. Harvey, M. J. Carman, I. Ruthven and F. Crestanig, "Bayesian latent variable models for collaborative item rating prediction", *Proc. ACM Conf. Inf. Knowl. Manage.*, pp. 699-708, 2011J. Noel, S. Sanner, K. N. Tran, P. Christen, L. Xie, E. V. Bonilla, E. Abbasnejad and E. D. Penna, "New objective functions for social collaborative filtering", *Proc. 21st Int. Conf. World Wide Web*, pp. 859-868, 2012
13. H. Ma, H. Yang, M. R. Lyu and I. King, "SoRec: Social recommendation using probabilistic matrix factorization", *Proc. ACM Conf. Inf. Knowl. Manage.*, pp. 931-940, 2008 [CrossRef]
14. M. Jiang, P. Cui, R. Liu, F. Wang, W. Zhu and S. Yang, "Scalable recommendation with social contextual information", *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 11, pp. 2789-2802, 2014
15. P. Winoto and T. Tang, "If you like the devil wears prada the book, will you also enjoy the devil wears prada the movie? a study of cross-domain recommendations", *New Generation Comput.*, vol. 26, no. 3, pp. 209-225, 2008
16. J. Tang, S. Wu, J. Sun and H. Su, "Cross-domain collaboration recommendation", *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, pp. 1285-1293, 2012 [CrossRef]
17. C. Li and S. Lin, "Matching users and items across domains to improve the recommendation quality", *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, pp. 801-810, 2014
18. Y. Shi, M. Larson and A. Hanjalic, "Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges", *ACM Comput. Surv.*, vol. 47, no. 1, pp. 3, 2014 [CrossRef]
19. S. Berkovsky, T. Kuflik and F. Ricci, "Cross-domain mediation in collaborative filtering", *Proc. 11th Int. Conf. User Model.*, pp. 355-359, 2007
20. S. Gao, H. Luo, D. Chen, S. Li, P. Gallinari and J. Guo, "Cross-domain recommendation



- via cluster-level latent factor model,"*Proc. Mach. Learn. Knowl. Discovery Databases*, pp. 161-176, 2013
21. W. Chen, W. Hsu and M. L. Lee, "Making recommendations from multiple domains,"*Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, pp. 892-900, 2013
 22. J. Tang, X. Hu, H. Gao and H. Liu, "Exploiting local and global social context for recommendation,"*Proc. 23rd Int. Joint Conf. Artif. Intell.*, pp. 2712-2718, 2013
 23. J. Tang, X. Hu and H. Liu, "Social recommendation: A review,"*Soc. Netw. Anal. Mining*, vol. 3, no. 4, pp. 1113-1133, 2013 [[CrossRef](#)]
 24. S. Sedhain, S. Sanner, L. Xie, R. Kidd, K. N. Tran and P. Christen, "Social affinity filtering: Recommendation through fine-grained analysis of user interactions and activities,"*Proc. 1st ACM Conf. Online Soc. Netw.*, pp. 51-62, 2013 [[CrossRef](#)]
 25. B. Shapira, L. Rokach and S. Freilikhman, "Facebook single and cross domain data for recommendation systems,"*User Model. User-Adapted Interaction*, vol. 23, no. 2-3, pp. 211-247, 2013
 26. S. Sedhain, S. Sanner, D. Braziunas, L. Xie and J. Christensen, "Social collaborative filtering for cold-start recommendations.", *Proc. ACM Conf. Recommender Syst.*, pp. 345-348, 2014
 27. Y. Chen, Y. Lin, Y. Shen and S. Lin, "A modified random walk framework for handling negative ratings and generating explanations,"*ACM Trans. Intell. Syst. Technol.*, vol. 4, no. 1, pp. 12, 2013
 28. G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions,"*IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734-749, 2005
 29. S. J. Pan and Q. Yang, "A survey on transfer learning,"*IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345-1359, 2010
 30. B. Cao, N. N. Liu and Q. Yang, "Transfer domains,"*Proc. Int. Conf. Mach. Learn.*, pp. 159-166, 2010
 31. Y. Zhu, Y. Chen, Z. Lu, S. J. Pan, G.-R. Xue, Y. Yu and Q. Yang, "Heterogeneous transfer learning for image classification,"*Proc. 25th AAAI Conf. Artif. Intell.*, pp. 1-9, 2011
 32. O. Moreno, B. Shapira, L. Rokach and G. Shani, "Talmud: Transfer learning for multiple domains,"*Proc. ACM Conf. Inf. Knowl. Manage.*, pp. 425-434, 2012
 33. Z. Lu, W. Pan, E. W. Xiang, Q. Yang, L. Zhao and E. Zhong, "Selective transfer learning for cross domain recommendation,"*Proc. 13th SIAM Int. Conf. Data Mining*, pp. 641-649, 2013
 34. B. Tan, E. Zhong, M. K. Ng and Q. Yang, "Mixed-transfer: Transfer learning over mixed graphs,"*Proc. 14th SIAM Int. Conf. Data Mining*, pp. 208-216, 2014 [[CrossRef](#)]
 35. W. Pan, E. W. Xiang, N. N. Liu and Q. Yang, "Transfer learning in collaborative filtering for sparsity reduction,"*Proc. 25th AAAI Conf. Artif. Intell.*, pp. 230-235, 2010
 36. B. Li, Q. Yang and X. Xue, "Can movies and books collaborate? cross-domain collaborative filtering for sparsity reduction,"*Proc. 21st Int. Jont Conf. Artif. Intell.*, vol. 9, pp. 2052-2057, 2009
 37. H. Jing, A.-C. Liang, S.-D. Lin and Y. Tsao, "A transfer probabilistic collective factorization model to handle sparse data in collaborative filtering,"*Proc. Int. Conf. Data Mining*, pp. 250-259, 2014
 38. H. Tong, C. Faloutsos and J. Pan, "Fast random walk with restart and its applications,"*Proc. Int. Conf. Data Mining*, pp. 613-622, 2006
 39. M. Gori and A. Pucci, "ItemRank: A random-walk based scoring algorithm for recommender engines,"*Proc. Int. Jont Conf. Artif. Intell.*, pp. 2766-2771, 2007
 40. B. Gao, T. Liu, X. Zheng, Q. Cheng and W. Ma, "Consistent bipartite graph co-partitioning for star-structured high-order heterogeneous data co-clustering,"*Proc. ACM SIGKDD Int.*



Conf. Knowl. Discovery Data Mining, pp. 41-50, 2005

41. B. Gao, T. Liu and W. Ma, "Star-structured high-order heterogeneous data co-clustering based on consistent information theory," *Proc. Int. Conf. Data Mining*, pp. 880-884, 2006
42. C. Avin and B. Krishnamachari, "The power of choice in random walks: An empirical study," *Comput. Netw.*, vol. 52, no. 1, pp. 44-60, 2008
43. I. Safro, P. D. Hovland, J. Shin and M. M. Strout, "Improving random walk performance," *Proc. Int. Conf. Sci. Comput.*, pp. 108-112, 2009