# Statement of Objectives

I am a scholarly student, originally studied hard and demonstrated a distinguished talent in science when I am young. After 12 years' primary study, I finally got the admission to get a bachelor degree of Computer Science at Dalian University of Technology (DUT). During my undergraduate study, I gained the opportunity to be introduced to several interesting subjects, through the study of which I laid a sound and comprehensive foundation. Through the years, I have also been conferred a variety of scholarships which epitomize my prominence and outstanding academic performance. For instance, one of them was National Scholarship, the highest college student reward since only two most excellent students out of 150 were granted that glory at the moment. Thereafter I got an admission to the postgraduate program of our university with exemption from the entrance examination being one of the top 10% undergraduate student candidates in our Department of Network Engineering. Moreover, I was awarded a scholarship by the Government of China to pursue my master degree in Computer Science.

As a graduate student in Dalian University of Technology, I have done much work on research, especially on Recommendation Technology. In 2010, I took part in the research project named High-Confidence Wireless Communication Protocols for Complex Cyber-Physical Systems, which is founded by National Natural Science Foundation of China (NSFC). I have to say that it was by no means easy to design a new MAC protocol at that time, however, through my diligent attitude, extensive reading and self-learning in graduate-level courses, I successfully tackled all the difficulties. In this regard, I proposed an innovative MAC Protocol for Wireless Networks and published one paper in *IEEE CYBER 2012* conference. In addition, my thesis, which was based on the project, secured the highest score of 95/100 in that group I was in.

From 2011, I put myself in the research of Personalized Academic Collaborative Recommendation Algorithm. Currently I have proposed an academic random walk model for most valuable collaborators recommendation (MVCWalker), which is capable of providing differentiated potentials for the particular researcher with different cooperative priorities. This resulted in a dramatic improvement of recommendation precision, recall, and coverage in coauthored networks, compared with the state-of-art strategies. Based on my previous work, I have written two papers being the primary student author. One paper titled ACRec: A Co-authorship based Random Walk Model for Academic Collaboration Recommendation is accepted to publish on WWW'14 Companion Proceedings (by ACM Digital Library). The other paper is now under review by *IEEE Transactions on Emerging Topics in Computing*. Although there were trials and tribulations in the arduous path ahead, I decisively pressed ahead which later proved to be rewarding and worthwhile.

In addition to research work, I have worked on numerous projects (individual and jointly), granting me the first-hand experience of implementation including the noteworthy research and development projects for: Talent Teaching Assistant System (built using the Dot Net Framework, C#), Conference Walkman: Put the Meeting into Pocket (built using Python). I also have sought to apply that which I have accumulated into practical use. I applied for a software development internship in Mitalk SNS Group of Xiaomi Inc., a company like Apple in Chinese smartphone market. During my internship, I try my utmost to tackle complicated problems by independent thinking. Under the concentrated guidance by skilled professionals of the company and other outstanding engineers' assistance, I acquired a firm

grasp of machine learning theory as well as programming techniques.

For I have both solid research backgrounds with a sound programming ability, I do not intend to retire from studies without securing a doctorate in the desired area of study of Computer Science. I believe that my training in pursuit of Masters of Recommendation Technology and Smart Systems has not only greatly enhanced my logic analysis and algorithm designing abilities and without a doubt has had a profound impact on my research and inter-communication skills. I, therefore regard myself, as being highly interested in Recommenders and Smart Things, as they have been my major research area to this date as well hot research spots within Machine Learning.

In the light of all my experiences, my dedication to the distributed computing field, and my enduring passion for knowledge, I have thus chosen to pursue the pioneering Research Degree: Computer Science in your University. The PhD program at UCL possesses unparalleled strength and is renown in this field. Therefore, I sincerely hope I will have the privilege to further my research work in the doctoral program so as to realize my goals. If I have the honor to be admitted, I will not only work hard on my own major projects, but also take full advantage of the resource there and learn more from others to become more sensitive about effective research. I believe after several years' study, I will be regarded as a distinguish PhD student and get closer to my objective to be an outstanding researcher. After graduation of PhD, I will consistently conduct research work no matter in a university or in a research institute aiming at making a difference to the world by research and practice. I am confident that, if given further opportunity, I can create even more glorious progress and discover broader territory which will eventually benefit the whole society. I sincerely hope you could consider my admission to your program as a doctoral student, helping a girl realizing her dream.

# Research Statement

## 1 Background and Problem Statement

We are witnessing an epoch-making proliferation of information available on the Internet. Meanwhile, information overload prevents users from acquiring relevant information. To tackle this problem, the emergence of recommendation systems and techniques is now playing a critical part of the contemporary internet world by bringing people closer to the resources they really need.

In the context of academic social networks, researchers approach big scholarly data problems such as how to obtain information pertaining to a valuable collaborator. Previous studies have confirmed that researchers or research groups with well-connected cooperation networks tend to be more prolific as well as productive scholars are more inclined to cooperation [1]. Thus, it is imperative and vital for researchers to get acquainted with new valuable collaborators in academic social networks [2]. To satisfy these kinds of demands, for example making friends in academic social networks, many methods have been proposed to suggest potential new links, including link prediction and link recommendation [3, 4].

A social network (SN) can be described in an abstract way, as a graph of nodes (such as users or groups) that have certain relationships, for example, friendship and co-authorship. In traditional SN, link prediction and link recommendation methods have made significant improvements of link creations. A feature in SN called "People You May Know" has proved to be of merit in recommending friends based on a FOF (friend of friends) method. Besides, typical SN systems usually recommend friends that the users already know offline based on social relationship [5]. However, in the academic context, social relationship has a different meaning considering some academic backgrounds (e.g., research interests, co-authoring information, and academic reputation). Therefore, recommending researchers in academic social networks (based on big scholarly data) is an increasingly important topic.

My focus in this research proposal is on Academic Social Networks (ASN) where social links are formed by certain academic ties. For instance, two researchers from the same research institute will be connected for the social tie of same affiliation. Within ASN, a co-authorship social network is an extraordinary social network due to the academic property of co-authorship, which is a simple graph evolving from the author-paper binary graph. In the research world, suggesting new links can motivate researchers to build new collaborative relationships when they consider writing a new paper, and then help them acquire papers of high quality. Since co-authorship social networks are of intrinsic collaboration related values, I want to present an academic recommendation model for collaboration accordingly.

## 2 Related Work

Social networks have been studied for decades in an effort to comprehend the relationships between people and detect patterns in such interactions. Recently much research work has been done on how to utilize social network information to improve recommender systems [6]. For instance, Ma et al. [7] elaborated on how the incorporation of social network information is beneficial in improving recommender systems. Perugini et al. [8] suggested that recommendation has an intrinsic social element that is intended to connect people. In contrast to previous work in traditional friend recommendation field, there is also some other research work on collaboration recommendation. Lopes et al. [9] considered the researcher’s

publications area and the vector space model to generate collaboration recommendation in academic social network. Besides, co-authorship social networks analysis has been studied for a long time, of which there are some positive outcomes in terms of collaboration recommendation.

Considering combining the network structure with the features of nodes [10], some methods (for example link prediction and link recommendation) based on Random Walk with Restart (RWR) have been proposed in the state-of-the-art literature. RWR provided a good way to measure how close related two nodes are in a graph [11]. In [12], Mohsen et al. proposed a random walk model which combined the trust-based and collaborative filtering approaches for recommendation.

However, current research either lacks combining the network structure with the features of nodes, or treats the links with equal importance in RWR, neglecting whether the relationship is strong or not. To tackle these drawbacks, we present an academic random walk model for collaboration recommendation in co-authorship social networks.

## 3 Aim of proposal

The purpose of my study is to take advantages of users' research interests and scholars' collaborative relationships to design a collaboration recommendation strategy and a related system that is a fabulous conference social application for attendees. To achieve the goal of our study, we should carry on research on the following questions:

1) How to define users' research interests and how can they affect the collaboration relationship over time? Are there any available features of interests taking critical roles for the research?
2) What are the state-of-art collaboration recommendation algorithms and systems? Are there any valuable schemes to be adopted?
3) What kinds of issues need to be tackled when designing an application? How to organize its function modules to make it of keenly favorite UE ?
4) How to evaluate the performance of the recommendation algorithm and the system? What are the up-to-date metrics or tools or data sets? Are there any differences between off-line experiments and real traces?

## 4 Research Progress and Plan

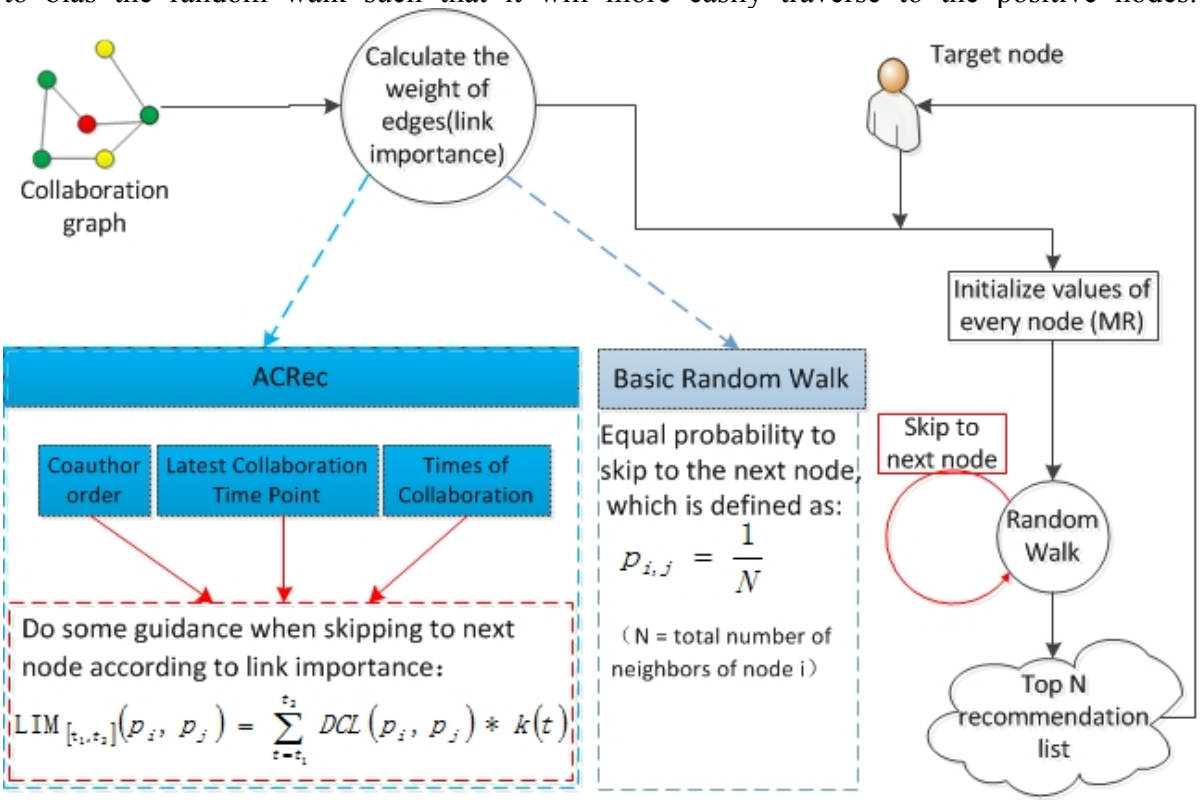
### 4.1 A Co-authorship based Random Walk Model for Academic Collaboration Recommendation

To satisfy the demand of collaboration recommendation through co-authorship in an academic network, we propose a random walk model using three academic metrics as basics for recommending new collaborations. Each metric is studied through mutual paper co-authoring information and serves to compute the link importance such that a random walker is more likely to visit the valuable nodes. Our experiments on DBLP dataset show that our approach can improve the precision, recall rate and coverage rate of recommendation, compared with other state-of-the-art approaches.

4.1.1 Academic Collaboration Recommendation Model

The academic collaboration recommendation model is inspired by the truth that scholars usually desire to cooperate with people who have high academic value. Such people normally have fruitful high-quality papers, which can generally be used to represent people's

academic achievements. Besides, as the RWR model has been proved to be competent for calculating the similarity of nodes in network, we use it as a basic model for the co-authorship social networks. And the three metrics we introduced into the network structure is to bias the random walk such that it will more easily traverse to the positive nodes.




The structure of Academic RWR is depicted in figure above. The whole collaboration recommendation work can be called as ACRec. Firstly, we extract a collaboration graph based on volumes of co-authored paper information. Then, as part of our major contributions, we derive three academic metrics from large scholarly data and then take them to compute each edge strength (link importance), which is going to be utilized to guide a co-authorship based random walk model (also called as Academic RWR). When Academic RWR ends, we can generate a top N recommendation list.

4.1.2 Three Academic Metrics

The three co-authorship metrics those are co-author order, latest collaboration time point, and times of collaboration respectively.

1) Co-author Order: In most cases, there is a list of co-authors for one paper. Normally, their contributions to the paper differ from each other. For example, the first and the second authors usually make more contributions than the rest authors. In such cases, the cooperation relationship between the first two authors is competently strong. Moreover, the co-author order can reflect cooperation relationship strength. As a general rule, the contribution value is inversely proportional to the co-author order, and the relevant two nodes contribute the weight of relationship. Therefore, we propose a measure of the link importance based on the coauthor order: DCL (distance in coauthor list). However, no previous study has taken into consideration the co-authoring order.
2) Latest Collaboration Time Point: Academic social networks and scholars' interests are time-varying, where the links among scholars change over time. For instance,

scholars might be more willing to collaborate with who they co-authored a paper last month, as compared to the coauthors that they cooperated ten years ago. Hence, we measure the link dynamics.

3) Times of collaboration: In academic social networks, if two authors coauthor a paper, there will be a link between them. Furthermore, these two authors may collaborate many times.

#### 4.1.3 Experiments

The essential experiments include:

1) Analyzing the data from big DBLP dataset to derive the co-authoring relationship among authors and their research interests.
2) Designing experiments for algorithms evaluation.
    a) We can divide the dataset into two parts: the data before year 2011 as a training set, and others as a testing set.
    b) We designed different experiments to compare ACRec with the basic model of RWR.
    c) The metrics we use to evaluate the performance are precision, recall rate and coverage rate.
    d) The parameters we consider include range of target nodes' degree, damping coefficient and the number of recommended nodes.

### 4.2 A Smart Conference System

Based on the proposed collaboration recommendation algorithm, we can design a smart conference system, which will be a very useful and creative mobile application for conference attendees. The system will include three core parts: conference schedule management, instant academic chat platform and collaboration recommendation to support the former two components. The conference schedule management part can formulate a personalized schedule for every user. In accordance with the recommended schedule, users can meet researchers they want to communicate with, attend sessions they have interest in, acquire information they focus on. Then the academic chat platform can provide users an online interaction opportunity, which will enhance the real world collaborative activities. The kernel of the system is the recommendation part. It helps users easily meet valuable and interesting people and information from big complex data environment, which is like some kind of serendipity.

### 4.3 Further Discussion

It is a long way to have the entire proposal realized with the remaining problems to be resolved. Though The proposal has a nice and non-trivial idea, the weighing of edges to make better informed decisions does not seem to be the novel idea as several works use weighted networks in making recommendations. We use time information (how much past and how often the authors worked in the past). One novelty is the use of the author order to weigh the edges; capturing the intuition that authors close-by in the author list are more related to each other. However how these metrics are combined is somewhat arbitrary, thus there remains quite an important task to figure out the weighting problem by massive experiments and discussion.

The evaluation is based on DBLP. Considering the data set is large and sparse, there is a need to do a good job in trying to find the best set of RWR parameters before comparing

their model with the basic RWR. As a consequence, we can just precisely see how significant is the improvement and the proposed weighing significantly better.

In the same time, the smart conference system is indeed an on-going program. So, there's no doubt about putting the system to be active.

## References


[1] S. Lee and B. Bozeman. The impact of research collaboration on scientific productivity. Social Studies of Science, 35(5): 673–702, 2005.

[2] H.-H. Chen, L. Gou, X. Zhang, and C. L. Giles. Collabseer: A search engine for collaboration discovery. In Proc. of ACM/IEEE JCDL'11, pages 231–240, Ottawa, Ontario, Canada, 2011.

[3] J. Tang, S. Wu, J. Sun, and H. Su. Cross-domain collaboration recommendation. In Proc. of ACM KDD'12, pages 1285–1293, Beijing, China, 2012.

[4] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su. Arnetminer: extraction and mining of academic social networks. In Proc. of ACM KDD'08, pages 990–998, Las Vegas, Nevada, USA, 2008. ACM.

[5] D. M. Boyd and N. B. Ellison. Social network sites: Definition, history, and scholarship. Journal of Computer-Mediated Communication, 13(1): 210–230, 2007.

[6] H.-H. Chen, L. Gou, X. Zhang, and C. L. Giles. Collabseer: A search engine for collaboration discovery. In Proc. of ACM/IEEE JCDL'11, pages 231–240, Ottawa, Ontario, Canada, 2011.

[7] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King. Recommender systems with social regularization. In Proc. of ACM WSDM'11, pages 287–296, Hong Kong, China, 2011.

[8] S. Perugini, M. A. Gon ̧calves, and E. A. Fox. Recommender systems research: A connection-centric survey. Journal of Intelligent Information Systems, 23(2):107–143, 2004.

[9] G. Lopes, M. M. Moro, L. K. Wives, and J. P. M.de Oliveira. Collaboration recommendation on academic social networks. In Advances in Conceptual Modeling ́lC Applications and Challenges, volume 6413 of Lecture Notes in Computer Science, pages 190–199. Springer Berlin Heidelberg, 2010.

[10] L. Backstrom and J. Leskovec. Supervised random walks: Predicting and recommending links in social networks. In Proc. of ACM WSDM'11, pages 635–644, Hong Kong, China, 2011.

[11] H. Tong, C. Faloutsos, and J.-Y. Pan. Random walk with restart: Fast solutions and applications. Knowledge and Information Systems, 14(3): 327–346, 2008.

[12] M. Jamali and M. Ester. Trustwalker: A random walk model for combining trust-based and item-based recommendation. In Proc. of ACM KDD'09, pages 397–406, Paris, France, 2009.