Recommending potential research collaborations based on science map using link prediction

XiaoMei Wang*, Tao Han and Ting Chen

*wangxm@mail.las.ac.cn
National Science Library, Chinese Academy of Sciences, Beijing (China)

Introduction

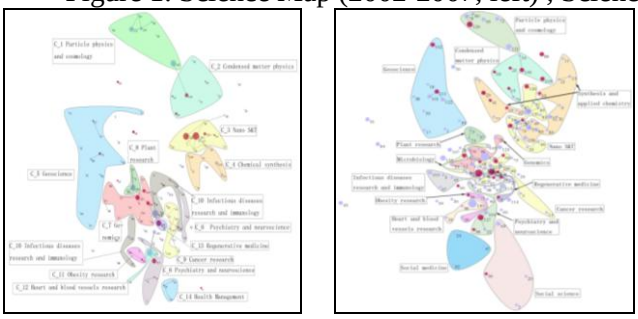
Research collaboration is an important topic. This study uses link prediction to recommend potential research collaborations in hot research areas of science map.

Link prediction aims at estimating the likelihood of the existence of links between nodes. After comparing 32 predictors of link prediction to examine which predictor is more accurate in the context of science map. We use the more accurate predictor to look for opportunities for future collaboration in research institution collaboration networks. A case study for Chinese Academy of Sciences (CAS) is present. It is expected that CAS scientists could identify the most fruitful targets for future collaboration in hot research areas.

Previous Work: Science Map and Research Area

Two period global Science Map were generated and hot research areas were identified as Figure 1. The Science Map's data is from top 1% highly cited papers in ESI. First research fronts are created by co-citation clustering of highly cited paters. Then hot research areas are also created by co-citation clustering of research fronts and then form a science map visualized by Gravity Model mapping (a force-directed method) at the level of research area.

Figure 1. Science Map (2002-2007, left) , Science Map (2004-2009, right) and their properties



Data	Science Map (2002-2007)	Science Map (2004-2009)
Highly cited papers	18 203	16 397
Research fronts	2 300	2 094
Hot research areas	121	132

Methods

Institution collaboration *network* Weight networks of co-authorships at institution level in hot research areas are *constructed* respectively in two period.

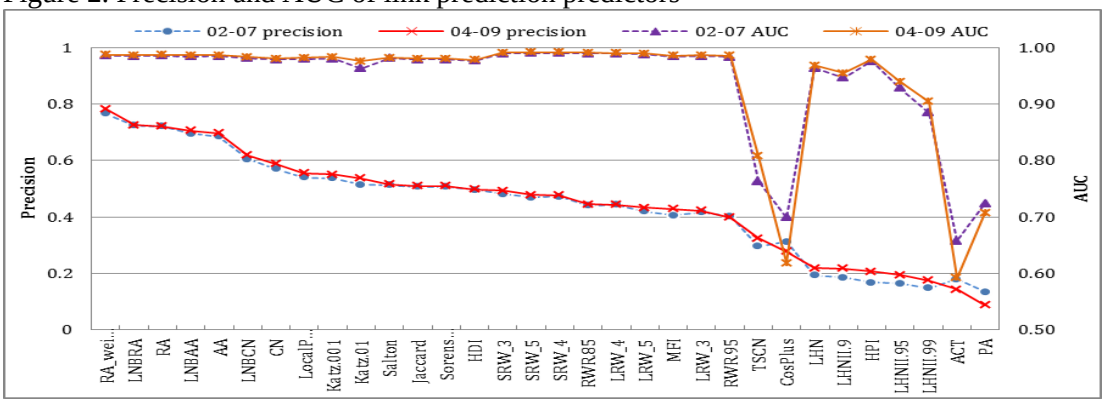
Link prediction techniques Every network is split into two parts for trainning(90%) and testing(10%), then 32 similarity-based predictors of link prediction are compared to examine which predictor is more accurate in the context of science map. Predictors include local-information-based similarity, such as Common Neighbors(CN), Salton Index, Jaccard Index, Hub Promoted Index(HPI), Leicht-Holme-Newman Index(LHNI), Preferential Attachment Index(PA), Resource Allocation Index(RA), path-based-similarity, such as Local Path Index(LP), Katz Index, random-walk-based similarity, such as Random Walk with Restart(RWR),SimRank, Local Random Walk(LRW),Superposed Random Walk(SRW) etc.

Recommendations for CAS Generally researchers working on the same topic would collaborate more possibly. Therefore, top 5 recommendation organizations in every hot research area are selected by two criterions: 1) Top ranked in the list of link predictor's recommendation. 2) Has co-occurrence (not co-authorship) relationship with CAS in one research front.

Recommendation result

Evaluation of link prediction algorithms Thirty-two predictors have their own different underlying philosophies and yield different accuracy in respect of two index, precision and AUC. Every predictor in every research area and mean value of all research areas was calculated in two periods science map. As showed in Figure 2, all predictor's mean value are highly consistent in both two science maps, it is means that the predictors are stable in the context of science map and the method could be effect. Among the 32 predictors, *weighted Resource Allocation* predictor has the highest precision and the almost highest AUC. So we could choose *weighted Resource Allocation* predictor in the following recommendation.

Figure 2. Precision and AUC of link prediction predictors



The case recommendations in some hot research area for CAS are listed in Table 1.

Table 1. Example of collaboration recommendations for CAS in Science Map 2004-2009

Rank	Hot Research Area (Science Map 2004-2009)				
	Standard cosmological model, dark energy and the relativistic theory of gravity (RA ID: 73)	Application of gold nanorods / particles, carbon nanotubes, fullerenes, quantum dots and metal oxide in biomedicine (RA ID: 118)	QCD phase transition and heavy ion collisions (RA ID: 121)	Theory and application of graphene (RA ID: 126)	Micro RNA research (RA ID: 129)
1	University of Toronto	University of California, Berkeley	Polytechnic University of Turin	Institute of Physics ASCR	National Taiwan University
2	Johns Hopkins University	Sungkyunkwan University	State University of New York at Stony Brook	Hitachi Cambridge Laboratory	EMMS Corporation
3	University of Oxford	Institute of Materials Research and Engineering	Joint Institute for Nuclear Research	National Physical Laboratory	National Institutes Of Health
4	University of California, Berkeley	National University of Singapore	Charles University in Prague	University of Wurzburg	Thomas Jefferson University
5	University of Washington	Johns Hopkins University	CERN	University of California, Santa Barbara	Nanjing Agricultural University

Conclusion

A possible approach was provided to suggest research collaborations based on science map. The accuracy from the perspective of predicted is not high when testing in continues research area of different period. Following we should combine the index and attribute of research area to improve the accuracy. Furthermore, machine learning algorithm could be considered.

References

Pan, JF., Zhang, XL., Wang, XM., etc. (2010). Mapping Science Structure 2009. Science Press.

Lu, LY & Zhou, T. (2011). Link precision in complex networks:A survey. Physica A, 390, 1150-1170

Boyack, KW. (2009). Using Detailed Maps of Science to Identify Potential Collaborations. Scientometrics, 79(1), 27-44.