

COSNET: CONNECTING HETEROGENEOUS SOCIAL NETWORKS WITH LOCAL AND GLOBAL CONSISTENCY

YUTAO ZHANG^{*}, JIE TANG^{*}, ZHILIN YANG^{*}, JIAN PEI[‡], AND PHILIP S. YU^{*†}

*Tsinghua University, [#]Simon Fraser University, [†]University of Illinois at Chicago

AMINER - ACADEMIC SOCIAL NETWORK ANALYSIS AND MINING SYSTEM





Integrating AMiner academic social network with Google Scholar, Linkedin, USPTO and Videolectures.

LOCAL, NETWORK AND GLOBAL CONSISTENCY

Local Matching: We formalize the similarity of user profiles by defining a set of local features for a single user pair x_i . Network Matching: We consider "neighborhood preserving matching". If user $v_i^1 \in G^1$ is matched onto $v_i^2 \in G^2$, then we also hope that v_i^1 's neighbors can be also matched to v_i^2 's



neighbors in G^2 .

Global Consistency: Given three networks G^1 , G^2 and G^3 , connecting $\langle v_i^1, v_j^2 \rangle$ and $\langle v_i^2, v_k^3 \rangle$ without connecting $\langle v_k^3, v_i^1 \rangle$ will cause the inconsistency problem, because the connections did not consider the transitivity of user identities.

COSNET: THE PROPOSED MODEL

We define an energy function E(Y, X) to model the likelihood of a configuration of *X* and *Y*. The problem of connecting multiple social networks is to find an optimal configuration Y^* given the user pairs X:

 $Y^{\star} = \arg\min_{Y \in \mathcal{V}} E(Y, X)$



(b) The generated matching graph (c) Matching graph after pruning (a) Two input networks

(d) The constructed model

EXPERIMENTAL RESULTS

We perform experiments on two data collections: SNS and Academia. Each data collection consists of several social networks. In the different networks of each data collection, both errors the users and the meanings of the relationships in the different networks are very different.

COSNET achieves a 10 - 30% improvement over baseline methods.





Effects of Candidate Pruning: Propagation can significantly increase the coverage of candidates with a reasonable number of





Academia collection



SNS collection

Academia collection



Application Improvement: The external network information from LinkedIn can significantly improve the performance of expert finding (+5-10% by Precision@5).

Effects of Global Consistency: 78 Without considering global consistency, the overall performances $\hat{\mathbb{S}}_{72}^{74}$ drop significantly (-2% and -7%; ¹¹ 70 *p*-value $\ll 0.01$, *t*-test) on both data collections. #networks

ONLINE SYSTEM

The online system is available at http://aminer.org The dataset & code are available at http://aminer.org/cosnet