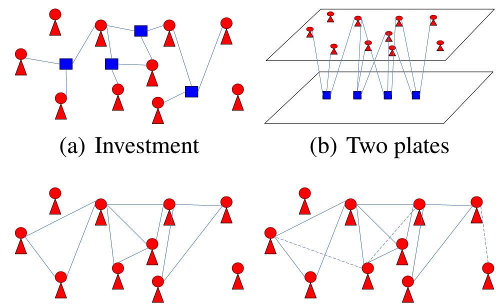
Who Will Be Your Next Co-investor?

Abstract

It is challenging to predict whether two venture capitals (VCs) will co-invest in the recent future or not. In this paper, we formulate the problem of co-investment prediction into a factor graph model incorporating structural balance theory. Experimental results demonstrate that the performance of the proposed model significantly (+14% in terms of accuracy) outperforms the baseline methods, such as logistic regression and SVM. In addition, we have some interesting findings, e.g., in VC network, the co-investor of my co-investor tends to be my co-investor; VCs with similar property right are likely to co-invest; Chinese investors are more liable to social relation than foreign investors.



(a) Investment (b) Two plates

(c) Synd. net (d) New co-inve.

Figure 1: Investment and co-investment.

Introduction

Social network plays an important role in economic action, and revealing the formation of social network has a significant meaning for both social network theory and economics. The co-investment of Venture Capitals (VCs) is an important economic event, which can also be regarded as a link in co-investment network. In this work, we study the problem of predicting whether two VCs will co-invest in the recent future or not. We address the challenges as follows. First, what are the fundamental factors that influence the formation of co-investment relationships? Second, how to design a mechanism that incorporate the social network theory affecting the formation of co-investment relationships.

Fig. 1 shows the investment and co-investment in the capital market. In Fig. 1(a), the red person represents a VC, the blue box represents an enterprise that gets fund, and a line between a VC and an enterprise represents an investment. VCs and enterprises are in different spaces (called heterogeneous network in computer science, or two modes network in sociology and economics), and we use two plates to indicate the different spaces, as shown in Fig. 1(b). To simplify the network, now we concern the syndication of VCs by adding a link between two VCs that have invested a common enterprise in the past, as shown in Fig. 1(c). Given the syndication network (defined in section 2) in time span $\{1, t\}$, we'd like to predict whether two VCs will co-investment not in time $t+1$, and the dashed lines in Fig. 1(d) are new co-investments in time $t+1$.

Co-investment has been studied for many years in sociology and economics, such as (Lerner 1994), (Sorenson and Stuart 2001) and (Kogut, Urso, and Walker 2007). However, most of the existing researches focus on the node attribute, and few works predict the future co-investment and present the performance of prediction.

Solution and contribution. In this paper, we formulate the prediction of co-investment in syndication network and perform a series of observations in the data. Based on the observations, we propose a structural balance based factor graph model named SBFG to predict the co-investment at time $t+1$ given syndication network of time span $(1, t)$. In the prediction, we not only consider topological network attributes in common practice, such as shortest distance, common neighbor and clustering coefficient, but also consider the attributes related to investment domain, such as property right and invested fields. We develop an approximate algorithm using loopy belief propagation to efficiently learn the proposed model. Our experimental results show that the proposed SBFG model can achieve a significant better performance (+14% in terms of accuracy) than the baseline methods.

Organization. Section 2 formulates the problem. Section 3 introduces the data set and our observations to verify the hypotheses. Section 4 proposes the structural balance based factor graph model and learning algorithm. Section 5 presents the experimental settings and results. Section 6 reviews the related work and section 7 concludes the paper.

Problem Formulation

In this section, we first present several definitions and then propose a formal description of the problem.

Definition 1 (Co-investment). We say that two VCs co-invest in a given year, if they invest the same enterprise(s) in the year.

Definition 2 (Syndication). We say that two VCs are syndicated, if they invested the same enterprise(s) in the past.

Note that co-investment only occurs when two VCs invest the same enterprise in the same year, while syndication occurs when two VCs invest the same enterprise in same year or different year. The number of investments increases over time, and the syndication network ($G^t = (V^t, E^t)$) are also evolving, where V^t is the set of $|V^t| = N^t$ VCs and $E^t \subseteq V^t \times V^t$ is the set of syndication relationships between VCs until time t. Let $\mathbf{w}_i$ be the set of attributes associated with VC v_i. An attribute can be VC's property right, invested fields, and so on. The attribute of VC could change over time, and we use $W^t = \mathbf{w}_1^t, \cdots, \mathbf{w}_N^t$ to denote the attributes of all VCs at time t. Our goal is to predict whether two VCs will co-invest or not in the next year, given their attributes and the existing syndication network. More specifically, we are concerned with the following problem.

Problem 1. Predict whether two VCs will co-invest or not in the next year. Let $G^t = (V^t, E^t, W^t)$ be the attribute augmented syndication network in time span $(1, t)$, given two VCs, the task is to predict whether they will co-invest or not in the next year (the period from t to t+1), which can be formulated as a binary classification problem.

Although researches in sociology and economics have investigated a large number of features that affect the co-investment, few of them present the prediction performance when these features are combined together. We need to explore the fundamental factors for high precision prediction of co-investment.

Data and Observation

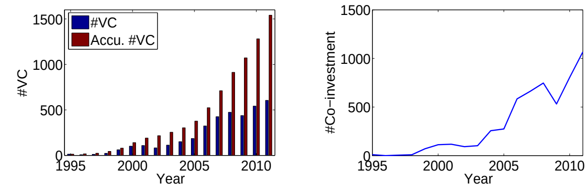
Data Collection

The data come from multiple resources, including two private databases Zero2IPO[1](Dec. 12th, 2012) and ChinaVenture[2], annual reports of China Venture Capital Research Institute[3], and national reports on "Venture Capital Development in China" by Chinese Academy of Science and Technology for Development. Moreover, we sample a subset from the data and conduct offline investigation to validate the truth of the data. The dataset contains open investment events from 1995 to 2011, there are totally 1541 VCs and 5455 co-investment events in the data. Fig. 2(a) and Fig. 2(b) show the rapid increase of VCs that invested in China and co-investments in the past 16 years (1995-2011). The number of co-investments today is nearly 100 times higher than 16 years before.

[1]http://www.zero2ipo.com.cn/

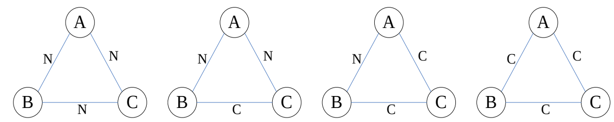
[2]http://www.chinaventure.com.cn/

[3]http://www.cvcri.com/



(a) #VC (b) #Co-investment

Figure 2: Number of VCs and co-investments over year



(a) (b) (c) (d)

Figure 3: Illustration of structural balance theory.

Observation

Before proposing our model, we first investigate the patterns that affect co-investment. We first explore the structural balance phenomenon in VC network, and then consider the patterns that closely related to capital behavior, including property right, invested fields complementarity/homogeneity and follow-the-trend. Finally, we present the patterns related to dynamic network, such as shortest distance, number of common neighbors, and clustering coefficient.

Structural balance. We connect our work to the structural balance theory(Easley and Kleinberg 2010), to see whether the co-investor of my co-investor is likely to be my co-investor. Fig. 3 shows the triad relationships, where C represents co-investor, and N represents non co-investor. For every group of three users (called triad), the structural balance theory implies that either all three pairs of these VCs are co-investors or only one pair of them are co-investor.

As shown in Fig. 4, the probability of balanced triads is by far larger than the number of unbalanced triads, which indicates that in capital market, the co-investor of my co-investor is likely to be my co-investor.

Distance of property right. According to property right theory(Grandori 2005), the property right (PR, also called capital type) of a VC can be categorized into four types by principal shareholders, i.e. CHInese VC, Chinese-Foreign VC (mixture PR), FOReign VC and OTHer. In general, VCs

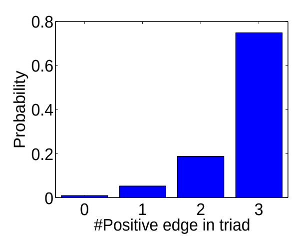


Figure 4: Structural balance. Y-axis: Probability, conditioned on the number of positive edges in the triad.

Table 1: The distance of property right

	CHI	C-F	FOR	OTH
CHI	0	1	2	1
C-F	1	0	1	1
FOR	2	1	0	2
OTH	1	1	2	0

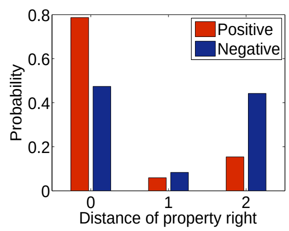
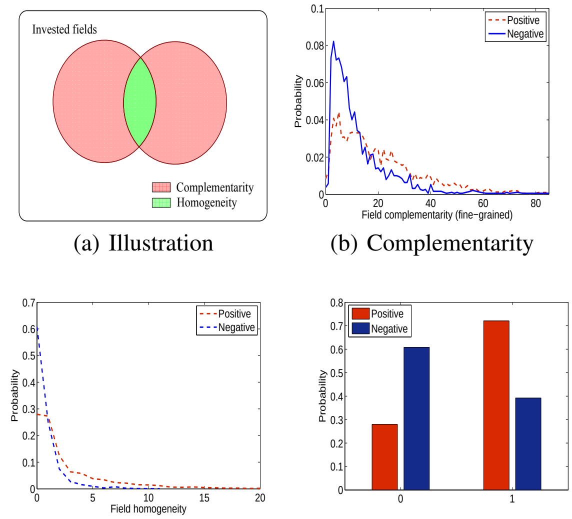


Figure 5: Distance of property right. Y-axis: probability density, conditioned on the distance of property right.

with similar property right are more likely to co-invest(Luo et al. 2014 Under review). To incorporate this prior domain knowledge, we define the distance of property right, as shown in Tab. 1, where 0 means there is no difference in PR between two VCs, 1 means there is a little difference between two VCs, and 2 means there is a large difference between two VCs. The probability conditioned on the distance of property right is shown in Fig. 5, where the orange bar represents positive instance in the dataset (described in section 5), and blue bar represents negative instance. The pattern clearly shows that the VC pair with no difference in PR is likely to co-invest, and the VC pair with large difference in PR is not likely to co-invest.

Invested fields complementarity/homogeneity. The investment preference is a set of special behavior of investment choice when individual investors confront risks and uncertainties, especially when they consider to co-invest. Both complementarity and homogeneity of invested fields will influence the co-investment between two VC firms. The invested fields are categorized into 20 coarse-grained types or 205 fine-grained types, and the fine-grained types are used in this paper. The complementarity of invested fields is defined as the symmetric difference of the sets of their invested fields, which is shown as the pink part in Fig. 6(a), while the homogeneity is defined as the intersection, which is shown as the green part in Fig. 6(a). There is another definition of homogeneity where 0 means the two VCs do not have common fields, and 1 means that the two VCs have common fields. As shown in Fig. 6(b), when two VCs complement each other to a large extent (complementarity is larger than 14), they tend to co-invest. Fig. 6(c) and 6(d) show that when two VCs have common fields, they tend to co-invest. When they do not have common field, they are not likely to co-invest, probably due to the lack of common interests.

Invested fields follow-the-trend. (Short-term) follow-the-trend refers to the "sheep-flock effect" that VC firms would choose to invest some hot fields to keep consistent with their peer, and the opportunism makes the investment behavior of VC barely according to the rational assumption(Christenson 1965)(Powell et al. 2005). The (short-term) follow-the-trend is defined as the ratio of number of invested fields of a given VC in the last year to that of all VCs in the last year, which reflects the irrational choice of VC. In order to reflect the overall preference of VC, we coin a new term, called long-term follow-the-trend, which is defined on the accumulated number of invested fields in the past. Fig. 7 shows the change of mean value of follow-the-trend over year[4]. Before 2002, the mean value of long-term follow-the-trend (long trend for short) was below that of short trend, which reflects the irrational choice of VCs in early capital market. Two curves became tangled in 2002-2004, while the long trend exceeded short trend after 2004, which means Chinese capital market became mature gradually.



(a) Illustration (b) Complementarity

(c) Homogeneity (def. 1) (d) Homogeneity (def. 2)

Figure 6: Invested field complementarity/homogeneity.

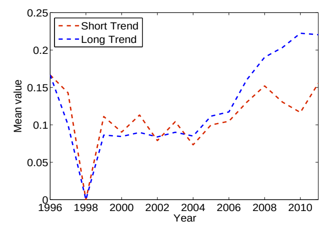
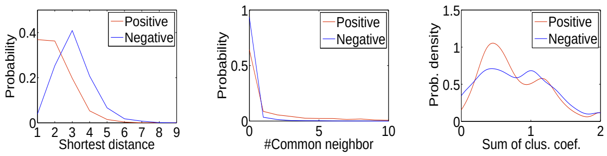


Figure 7: Turning point of follow-the-trend. Y-axis: mean value, conditioned on the year.

Besides the patterns presented above from domain knowledge, we also consider the topological network patterns, which are well-known in link prediction. Although there are many choices, we adopted shortest distance, number of common neighbors, and clustering coefficient.

Shortest distance. Shortest distance is considered to be one of the most important patterns in link prediction(Hasan et al. 2006). (Kleinberg 2000) discovered that in social net-

[4]The value 0 in 2008 is due to data sparseness.



(a) Shor. dist. (b) Comm. neig. (c) Clus. coef.

Figure 8: Topological patterns. Y-axis: probability density, conditioned on the number of shortest distance / #common neighbor / sum of clustering coefficient.

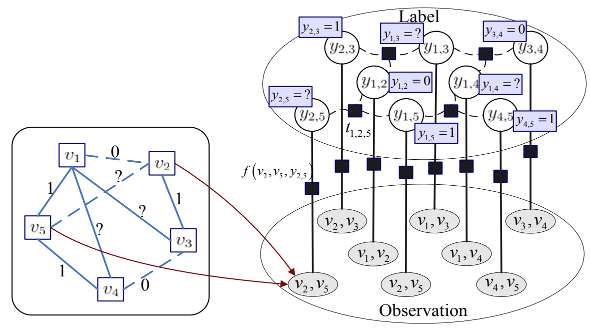


Figure 9: Graphical representation of SBFG model.

work most of the nodes are connected within a short distance, which makes it a very good pattern. As shown in Fig. 8(a), if the shortest distance between VCs in the syndication network is 1, i.e. they have co-invested before, they are very likely to co-invest again. If the shortest distance is 2, the two VCs still prefer to co-invest. When shortest distance is 3 or larger, they don't prefer to co-invest.

Common neighbors. We consider the link homophily of two VCs, which is based on the assumption that similar users tend to associate with each other. For the link homophily, we explore the number of common neighbors on co-investment. In Fig. 8(b), if there is no common neighbor, VCs don't prefer to co-investing, which is consistent with the situation where the distance of two VC is 3 or larger. If there are one or more common neighbors, VCs prefer to co-invest.

Clustering coefficient is considered to be an important pattern in social network research(Newman 2001). We used the sum of clustering coefficient of two VC as the pattern. As shown in Fig. 8(c), the VCs that in a proper dense neighborhood (the sum is between 0.2 and 0.8) is likely to co-invest, which may be probably explained by the coupling and decoupling theory(Granovetter 2002). In the theory, coupling is necessary to grasp the opportunity, but too tightly relationship keep the circle away from heterogeneous information, while decoupling can make the overly strong social network loosely open in this situation.

Model Framework

Basically, the binary classification problem (co-invest or not) can be solved by any classifier, such as logistic regression and SVM. However, these models suffer from the same limitation that they can not model the correlation between co-investments, but the correlation (structural balance) is very important in VC network. We model each possible co-investment as a node in a graphical model, and the problem is converted to predict whether the co-investment node takes value of 1 or 0 (co-invest or not).

The Proposed Model

Based on the above intuition, we propose a structural balance based factor graph model (SBFG) to predict whether two VCs will co-invest or not in the next year, and the graphical representation is shown in Fig. 9.

In Fig. 9, the left figure shows the original VC network, where the edges with label 1/0 represent whether two VCs co-invested or not in time span $\{1, t\}$, and the edges with label ? are those that we try to predict in time span $\{t, t+1\}$. The solid/dashed line means that the edge exists or not in the ground truth. The right figure is the SBFG model drived from the original VC network. $y_{i,j}$ is latent variable indicates whether two VCs v_1, v_2 co-invest. $f(v_i, v_j, y_{i,j})$ ($f_{i,j}$ for short) represents a attribute factor function defined for the co-investment. $t(y_{i,j}, y_{i,k}, y_{j,k})$ ($t_{i,j,k}$ for short) represents a triad factor function, which is used capture the structural balance among three possible co-investments $y_{i,j}, y_{i,k}, y_{j,k}$.

Y denotes the vector that contains all latent variables. Since we already know the co-investments in time span $\{1, t\}$, so the latent variables Y in the SBFG can be divided into labeled subset Y^L and unlabeled subset Y^U (to be predicted). For each latent variable $y_{i,j}$, we combine corresponding individual attributes $\{\mathbf{w}_i, \mathbf{w}_j\}$ and the derived aggregated attributes from individual attributed into a new attribute vector $\mathbf{x}_{i,j}$.

We formalize the network with Markov random fields. According to Hammersley-Clifford theorem(Hammersley and Clifford 1971), the probability of latent label Y given observations can be factorized as

$$p(Y|X) = \frac{1}{Z} \prod_{possible\ i,j} f_{i,j} \prod_{possible\ i,j,k} t_{i,j,k} \tag{1}$$

Where "$possible\ i,j$" means all possible values that i, j can take in the data set, and "$possible\ i,j,k$" has similar meaning. Factors are defined as

$$f_{i,j} = exp\{\boldsymbol{\alpha}_{i,j}^T \mathbf{g}(v_i, v_j, y_{i,j})\} \tag{2}$$

$$t_{i,j,k} = exp\{\boldsymbol{\beta}_{i,j,k}^T \mathbf{h}(y_{i,j}, y_{i,k}, y_{j,k})\} \tag{3}$$

Where $\mathbf{g}(v_i, v_j, y_{i,j})$ is the feature vector for attribute factor, $\mathbf{h}(y_{i,j}, y_{i,k}, y_{j,k})$ is the feature vector for triad factor, and $\alpha_{i,j}$, $\beta_{i,j,k}$ are corresponding weighting vectors. Furthermore, we pack all weighting vectors $\boldsymbol{\alpha}_{i,j}$, $\boldsymbol{\beta}_{i,j,k}$ into a long weighting vector $\boldsymbol{\theta}$, and pack all feature vectors $\mathbf{g}(v_i, v_j, y_{i,j})$, $\mathbf{h}(y_{i,j}, y_{i,k}, y_{j,k})$ into a long feature vector $\mathbf{s}$, regardless of the type of factors. Thus, the conditional probability, i.e. Eq. 1, is simplified to be

$$p(Y|X) = \frac{1}{Z} exp\{\boldsymbol{\theta}^T \mathbf{s}\} \tag{4}$$

Therefore, we try to get proper weighting vector $\boldsymbol{\theta}$ in the learning phase.

Learning

The latent variables in time span $\{1, t\}$, i.e. Y^L, are labeled, and our optimization goal is to maximize the log-likelihood of the labeled variables:

$$\begin{aligned} O(\boldsymbol{\theta}) &= logp(Y^L|X) = log \sum_{Y^U} p(Y^L, Y^U|X) \\ &= log \sum_{Y^U} exp\{\boldsymbol{\theta}^T \mathbf{s}\} - log \sum_{Y} exp\{\boldsymbol{\theta}^T \mathbf{s}\} \end{aligned} \tag{5}$$

To maximize the log-likelihood, we consider a gradient decent method, and the gradient is calculate as follows.

$$\frac{\partial O(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = E_{p(Y^U|Y^L,X)}[\boldsymbol{s}] - E_{p(Y^U,Y^L|X)}[\boldsymbol{s}] \tag{6}$$

One challenge here is that the graphical structure can be arbitrary and contain cycles, which makes it intractable to calculate the precise expectation, and we employ loopy belief propagation (LBP) (Frey and MacKay 1997) to approximate the expectation. Note that we should perform LBP twice in each step, one for estimating marginal probability $p(Y^U, Y^L|X)$, and the other for $p(Y^U|Y^L, X)$. In the end of each step, we update the weighting vector $\boldsymbol{\theta}$ with the gradient and an empirical learning rate η. The learning algorithm is shown in Algorithm. 1.

Algorithm 1: SBFG Learning Algorithm

Input: labeled variables Y^L, observations X, learning rate η
Output: weighting vector $\boldsymbol{\theta}$

Initialize $\boldsymbol{\theta}$;
while *not converged* **do**
  Calculate $E_{p(Y^U|Y^L,X)}[\boldsymbol{s}]$ using LBP;
  Calculate $E_{p(Y^U,Y^L|X)}[\boldsymbol{s}]$ using LBP;
  Calculate the gradient $\frac{\partial O(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ according to Eq. 6;
  Update $\boldsymbol{\theta}$ with $\boldsymbol{\theta}^{new} = \boldsymbol{\theta}^{old} - \eta \cdot \frac{\partial O(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$
Return $\boldsymbol{\theta}$;

Prediction

Once we get the learned weight vector $\boldsymbol{\theta}$, we can predict the unlabeled variable Y^U by first computing the marginal probability of $p(Y^U|Y^L, X)$ and then select the value with largest marginal probability as the label. Again, the marginal probability of $p(Y^U|Y^L, X)$ is calculated by running LBP, and the marginal probability is then taken as the prediction confidence.

Experiments

Experimental Setup

The dataset contains 1541 VCs that invested in China and 5455 co-investment events from 1995 to 2011. The 5455 co-investments are positive instances in our experiments. There are no direct negative instances in the dataset, and the we consider all possible combinations of accumulated VCs that have invested in China. However, the number of combinations is hundreds of times larger than the number of positive instances, and we sample the same number of negative instances as positive instances, which is a common practice in link prediction, such as (Hasan et al. 2006). Our goal is to predict co-investment in time $t + 1$ given data in time span $(1, t)$, and we construct two datasets. One is to predict co-investment in 2011 given data in 1995-2010, another is to predict co-investment in 2010 given data in 1995-2009.

Table 2: Prediction performance of co-investment

Data	Alg.	Pre.	Rec.	F1	Acc.
2011	SVC2	0.7075	0.4217	0.5285	0.6160
	SVC1	0.7233	0.5145	0.6013	0.6518
	LR2	0.7214	0.4733	0.5716	0.6380
	LR1	0.7311	0.4995	0.5935	0.6509
	FG	0.7400	0.5708	0.6444	0.6786
	SBFG	**0.8091**	**0.8819**	**0.8439**	**0.8336**
2010	SVC2	0.6740	0.6447	0.6590	0.6687
	SVC1	0.7134	0.5876	0.6444	0.6780
	LR2	0.6796	0.6298	0.6538	0.6687
	LR1	0.6969	0.6199	0.6561	0.6774
	FG	0.7149	0.6696	0.6915	0.7033
	SBFG	**0.7745**	**0.8236**	**0.7983**	**0.7933**
Average	SVC2	0.6907	0.5332	0.5938	0.6423
	SVC1	0.7184	0.5511	0.6228	0.6649
	LR2	0.7005	0.5515	0.6127	0.6533
	LR1	0.7140	0.5597	0.6248	0.6642
	FG	0.7274	0.6202	0.6680	0.6909
	SBFG	**0.7918**	**0.8528**	**0.8211**	**0.8135**

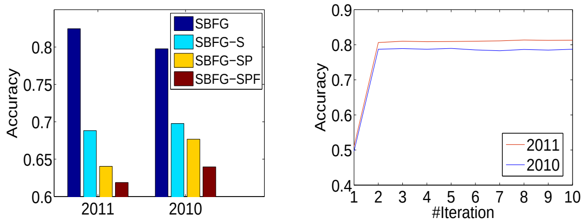
Prediction Performance

We compare our proposed model with other state-of-art supervised machine learning algorithms, and the results are shown in Tab. 2. The compared algorithm are (1)support vector classifier with L2 regularization and L2 loss function (SVC2), (2)SVC with L1 regularization and L2 loss function (SVC1), (3)logistic regression with L2 regularization (LR2), (4)LR with L1 regularization (LR1), and (5)SBFG model without social balance correlation (FG). The first four algorithms are implemented in LIBLINEAR software package(Fan et al. 2008), and the fifth algorithm is based on SBFG model by removing structural balance factor. As shown in Tab. 2, SBFG significantly exceeds all state-of-art algorithms in four measurements. The prediction accuracy and F1 of SBGF are around 0.8, which are satisfactory in co-investment prediction.

Analysis and Discussions

Factor contribution analysis. We examine the the contribution of different factors (patterns) by removing them one by one in the model. As shown in Fig. 10(a), SBFG-S excludes structural balance factor, SBFG-SP excludes both structural balance factor and property right factor, while SBFG-SPF further excludes factor related to invested fields (complementarity, homogeneity and follow-the-trend). When structural balance factor is removed from the model, the accuracy drops 10-13% in terms of accuracy. When all factor property right factor and invested fields factors are removed, there is only topological factors in the model, and the accuracy drops further 6-7%, which validates the effectiveness of proposed domain patterns in co-investment prediction.

Convergence analysis. We conduct experiments on the effect of the number of iterations of algorithm. Fig. 10(b) shows the convergence analysis results of learning algorithm. SBGF model converges very fast, usually within 10 iteration steps, which suggests that the learning algorithm is very efficient and has a good convergence property.



(a) Factor contribution (b) Convergence

Figure 10: Factor contribution analysis and convergence analysis.

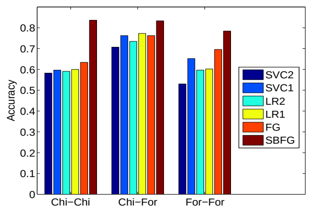


Figure 11: Accuracy on the country combination (2011). Y-axis: accuracy, conditioned on country combination.

Country analysis. We analyze the performance on VC pair with different country combination, i.e. Chinese-Chines, Chinese-Foreign and Foreign-Foreign. As shown in Fig. 11, in the case of Chinese-Chinese, SBFG exceeds other algorithms with a significantly larger margin than that of Chinese-Foreign or Foreign-Foreign, which suggests that Chinese investors are liable to social relation (structural balance), and they rely on the robustness of network to avoid risk.

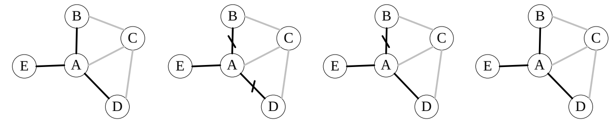
Qualitative Case Study

Now we present a case study to demonstrate the effectiveness of the proposed model. Fig. 12 shows a example generated from our experiments, where each node represents a VC (A is Alibaba Capital Partners, B is Crescent Point Group, C is Japan Asia Investment, D is Walden International Investment Group, and E is Sequoia Capital), gray line indicates that the two VCs co-invested before 2010, black line indicates that they co-invested in 2011, and the black line with a mark indicates that the algorithm makes a mistake. Our goal is to predict the co-investment in 2011 given data in 1995-2010. SVC2 correctly predicts the co-investment between A and E, but misses the other two ones. FG model predicts two co-investments, but still misses one. SBFG model correctly predicts all three co-investments, and this is because SBFG leverage the structural balance factor. After adding the co-investment between A and B, the triad ABC becomes a balanced one, and so is ACD.

Related Works

Co-investment and Syndication

In sociology and economics, the study of co-investment dates back to Wilson's theory on syndication(Wilson 1968), and (Lerner 1994) studied the principle of who will be a good co-investor and when to reconstruct a co-investment. More recently, some scholars studied the co-investment/syndication from the perspective of link formation, such as (Sorenson and Stuart 2001), (Piskorski 2004), (Kogut, Urso, and Walker 2007). Based on 45 years' VC data from U.S., (Kogut, Urso, and Walker 2007) found several attributes that might have influence on the new link. However, (Kogut, Urso, and Walker 2007) only used the node attributes and they did not build a predictor.



(a) Truth (b) SVC2 (c) FG (d) SBFG

Figure 12: Case study.

Link Prediction

Our work is closely related to link prediction, which is one of the fundamental tasks in social network. Existing work on link prediction can be broadly grouped into two categories based on the learning algorithms: unsupervised link prediction and supervised link prediction. The classic works of unsupervised prediction are surveyed in (Liben-Nowell and Kleinberg 2007) and recently (Lichtenwalter, Lussier, and Chawla 2010) designed a flow based method. There are a number of works on supervised link prediction, such as (Backstrom and Leskovec 2011)(Leskovec, Huttenlocher, and Kleinberg 2010)(Hopcroft, Lou, and Tang 2011)(Tang, Lou, and Kleinberg 2012)(Yang et al. 2012)(Wu, Sun, and Tang 2013). (Hopcroft, Lou, and Tang 2011) studies the extent to which the formation of a reciprocal relationship can be predicted in a dynamic network. (Tang, Lou, and Kleinberg 2012) developed a framework for classifying the type of social relationships by learning across heterogeneous networks. In this work, we focus on studying the underlying patterns that influence the formation of co-investment and propose a factor graph model to incorporate structural balance theory and discovered intuition.

Conclusion

In this paper, we study the prediction of co-investment of VC. We present a series of observation analysis and propose a factor graph model SBFG based on structural balance theory to formalize the observation into a unified model. For the model learning, we employ the loopy belief propagation to obtain an approximate solution. Experimental results show that the proposed method can accurately predict the co-investment in recent future, and obtains a significant improvement (+14% in terms of accuracy) over the baseline methods. In the future, we will further explore the prediction of co-investment in the heterogeneous network by incorporating the enterprises.

References

Backstrom, L., and Leskovec, J. 2011. Supervised random walks: Predicting and recommending links in social networks. In *WSDM'11*.

Christenson, C. 1965. *Strategic Aspects of competitive Bidding for Corporate Securities*. Hardvard University.

Easley, D., and Kleinberg, J. 2010. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press.

Fan, R.-E.; Chang, K.-W.; Hsieh, C.-J.; Wang, X.-R.; and Lin, C.-J. 2008. Liblinear: A library for large linear classification. *Journal of Machine Learning Research*.

Frey, B., and MacKay, D. 1997. A revolution: Belief propagation in graphs with cycles. In *NIPS'97*.

Grandori, A. 2005. Property right theory and governance models in strategy. *Journal of Management and Governance*.

Granovetter, M. 2002. A theoretical agenda for economic sociology. In *The New Economic sociology: Developments in an Emerging Field*, 35–59. Russell Sage Foundation.

Hammersley, J., and Clifford, P. 1971. Markov field on finite graphs and lattices. In *Unpublished manuscript*.

Hasan, M. A.; Chaoji, V.; Salem, S.; and Zaki, M. 2006. Link prediction using supervised learning. In *SDM'06*.

Hopcroft, J.; Lou, T.; and Tang, J. 2011. Who will follow you back? reciprocal relationship prediction. In *CIKM'11*.

Kleinberg, J. 2000. Navigation in a small world. *Nature*.

Kogut, B.; Urso, P.; and Walker, G. 2007. Emergent properties of a new financial market: American venture capital syndication, 1960-2005. *Unpublished manuscript*.

Lerner, J. 1994. The syndication of venture capital investments. *Financial Management*.

Leskovec, J.; Huttenlocher, D.; and Kleinberg, J. 2010. Predicting positive and negative links in online social networks. In *WWW'10*.

Liben-Nowell, D., and Kleinberg, J. 2007. The link-prediction problem for social networks. *JASIST*.

Lichtenwalter, R.; Lussier, J.; and Chawla, N. 2010. New perspectives and methods in link prediction. In *KDD'10*.

Luo, J.-D.; Zhou, L.; Zhou, Y.; and Tang, J. 2014, Under review. Why do chinese firms invest jointly? In *Annual Meeting of the Academy of Management*.

Newman, M. 2001. The structure of scientific collaboration networks. In *Proceedings of the National Academy of Sciences*.

Piskorski, M. 2004. Networks of power and status: Reciprocity in venture capital syndicates. *Unpublished manuscript*.

Powell, W.; White, D.; Koput, K.; and Owen-Smith, J. 2005. Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences. *American Journal of Sociology*.

Sorenson, O., and Stuart, T. 2001. Syndication networks and the spatial distribution of venture capital investment. *The American Journal of Sociology*.

Tang, J.; Lou, T.; and Kleinberg, J. 2012. Inferring social ties across heterogenous networks. In *WSDM'12*.

Wilson, R. 1968. The theory of syndicates. *Econometrica*.

Wu, S.; Sun, J.; and Tang, J. 2013. Patent partner recommendation in enterprise social networks. In *WSDM'13*.

Yang, Y.; Tang, J.; Keomany, J.; Zhao, Y.; Li, J.; Ding, Y.; Li, T.; and Wang, L. 2012. Ming competitive relationships by learning across hetergeneous networks. In *CIKM'12*.