The Prediction of Venture Capital Co-Investment Based on Structural Balance Theory

(Extended Abstract)

Yun Zhou, Zhiyuan Wang
State Key Laboratory of High Performance Computing
National University of Defense Technology
Changsha, Hunan 410073, P.R.China
Email: zy_vxd@nudt.edu.cn, wzy@nudt.edu.cn

Jie Tang, Jar-Der Luo
Tsinghua University
Beijing 100084, P.R.China
Email: jietang@tsinghua.edu.cn, jdluo@mail.tsinghua.edu.cn

***Keywords*—venture capital, co-investment, prediction, factor graph model, group Lasso.**

I. Introduction

Venture capital (VC) is financial capital provided to early-stage, high-potential startup firms. High-tech industry routinely acknowledges that communities knit together by networks of social relations are essential for the development of the industry, and emphasizes that VCs hold central positions in these networks [1]. Based on the statistics on the free online CrunchBase dataset (CRUNCH for short), 80.9% of VC investments are related to at least two investors, thus co-investment is an important phenomenon in the VC market.

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

We deal with the following problem.

Problem. Predict whether two VCs will co-invest or not in the next year. Let $G^t = (V^t, E^t)$ be the VC network in time span $\{1, t\}$, where V^t is the set of accumulated VCs, and $E^t \subseteq V^t \times V^t$ is the set of accumulated co-investment relationships between VCs until time t. Given G^t and two VCs, the task is to predict whether they will co-invest or not in time $t + 1$.

Due to complexity and uncertainty of VC behavior, it's challenging to accurately predict future co-investments, and we address the challenges as follows. First, what factors influence the formation of co-investment relationships? Second, how to select a small number of fundamental factors that best explain the formation? Third, how to design a mechanism that incorporates social network theory affecting the formation of co-investment relationships?

II. Feature Design, Feature Selection and Prediction Model

Most of existing researches on co-investments dealt with a small dataset and only explored a few features for co-investment without detailed analysis. In addition, few works predicted future co-investment with a unified model and presented the performance of prediction.

We design a lot of (81) features from the perspective of both VC domain knowledge and social network theory(cf. [2]), which cover most features proposed in past literature, such as [3], [4], [5] and [1]. In order to gain both interpretability of features and high accuracy, we select prominent features by group Lasso. It is shown that only the top 10 features selected by group Lasso can explain the formation of the VC network quite well, e.g. nationality, number of common neighbors, betweenness, shortest distance, investor type, number of invested fields and Jaccard index of invested fields.

For every group of three VCs (called triad), the structural balance theory[6] implies that either all three pairs of these VCs are co-investors or only one pair of them are co-investor. Based on structural balance theory and observations on the dataset, we propose a structural balanced factor graph model named SBFG to predict the co-investment at time $t+1$, given co-investment network of time span $\{1, t\}$.

In Fig. 1, the left figure shows the original VC network with VC as a node, where the edges with label 1/0 represent whether two VCs co-invest or not in time span $\{1, t\}$, and the edges with label ? are those that we try to predict in time $t+1$. However, it is hard to model the correlation between/among co-investments (e.g. structural balance phenomenon) if with VC as a node. Thus, the original VC network with VC as a node is converted to the SBFG model with co-investment as a node, which is shown in the right figure. $y_{i,j}$ is the latent variable that indicates whether two VCs v_i, v_j co-invest, and $x_{i,j}$ is the observation of two VCs v_i, v_j. $f_{i,j}$ is the feature factor for a co-investment, and $t_{i,j,k}$ is the triad factor for three possible co-investments.

We formalize the model with Markov random fields, and develop an approximate algorithm using loopy belief propagation to efficiently learn the proposed model (cf. [2] for details).

III. Experiments

We evaluate our model on CRUNCH[1], which contains open investment events in the world from 1984 to 2014,

Zhiyuan Wang is the corresponding author.

[1]http://www.crunchbase.com/, March 20th, 2014

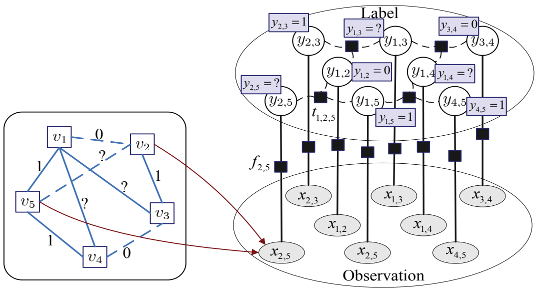

Fig. 1. Graphical representation of SBFG model.

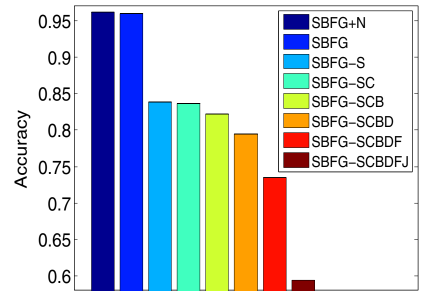

Fig. 2. Feature contribution analysis for the case of 2014.

TABLE I. PREDICTION PERFORMANCE OF CO-INVESTMENT WITH THE TOP 10 FEATURES.

Data	Alg.	Pre.	Rec.	F1	Acc.
2011	SVC	**0.8615**	0.7082	0.7773	0.8078
	LR	0.8601	0.7071	0.7761	0.8068
	SBFG	0.8236	**0.9939**	**0.9008**	**0.8963**
2012	SVC	**0.8770**	0.7059	0.7822	0.8129
	LR	0.8721	0.7095	0.7825	0.8122
	SBFG	0.8431	**0.9939**	**0.9123**	**0.9090**
2013	SVC	**0.8693**	0.7124	0.7831	0.8104
	LR	0.8664	0.7133	0.7825	0.8095
	SBFG	0.8395	**0.9920**	**0.9094**	**0.9050**
2014 (first 3 months)	SVC	0.9143	0.7210	0.8062	0.8287
	LR	0.9164	0.7240	0.8089	0.8309
	SBFG	**0.9308**	**0.9924**	**0.9606**	**0.9598**
Average	SVC	**0.8805**	0.7119	0.7872	0.8150
	LR	0.8788	0.7135	0.7875	0.8149
	SBFG	0.8593	**0.9931**	**0.9208**	**0.9175**

and there are a total of 18,716 VCs, 25,327 startups, 90,280 investments and 152,227 co-investments. We construct four cases for CRUNCH. The first case is to predict co-investments in 2011 given 1984-2010, the second is 2012 given 1984-2011, the third is 2013 given 1984-2012, and the fourth is the first 3 months in 2014 given 1984-2013.

We compare our proposed model (SBFG) with support vector classifier (SVC) and logistic regression (LR). Four measures, i.e. precision, recall, F1 measure and accuracy, are used to evaluate the performance. All algorithms use the top 10 features selected by group Lasso. As shown in Table I, SBFG significantly exceeds all state-of-the-art algorithms in all measures except precision.

We examine the contribution of different features by removing them one by one in the model for the case of 2014. As shown in Fig. 2, SBFG stands for the proposed method with the top 10 features. The plus mark denotes additional features besides the top 10 features, and the minus mark denotes features that are excluded from the top 10 features. N denotes the remaining 71 features other than the top 10 features, S the structural balance factor, C common neighbors, B betweenness, D shortest distance, F the number of invested fields, and J the Jaccard similarity of invested fields. When using only the top 10 features selected by group Lasso, the accuracy is 95.98%, which is only 0.18% less than 96.16% with a total of 81 features, which shows that the top 10 features can explain the formation of the VC network quite well. Note that, when the structural balance feature is removed from the model, the accuracy drops by 12.13% (from 95.98% to 83.85%), which demonstrates the prediction power of structural balance theory.

In addition, we have some interesting findings. For instance, in the VC network, the co-investor of my co-investor tends to be my co-investor; VC pairs from the same country, of the same investor type, with short distance, with more common neighbors or with appropriate Jaccard similarity of invested fields are likely to co-invest; VCs of large betweenness or of a large number of invested fields have advantage in the VC network; investors of Asian countries, especially of China, are more likely to have social relations than other countries.

IV. CONCLUSION

In this paper, we study the prediction of co-investment of VCs. We present a series of observation analysis, design a large number of features, and then select prominent features for co-investment by group Lasso. Then we propose a factor graph model SBFG based on structural balance theory to formalize the observation into a unified model. Experiment results show that the proposed method can accurately (around 90% in terms of accuracy) predict the co-investment in the near future with only 10 features selected by group Lasso, and obtains a significant improvement over the baselines.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China under Grant No. 61303068, and the Research Fund of State Key Laboratory of High Performance Computing under Grant No. 201502-02.

REFERENCES

[1] D. Trpido, "Mechanisms of venture capital co-investment networks: Evolution and performance implications," *Unpublished manuscript*, 2009.

[2] Z. Wang, Y. Zhou, J. Tang, and J.-D. Luo, "The prediction of venture capital co-investment based on structural balance theory," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 2, pp. 537–550, February 2016.

[3] O. Sorenson and T. Stuart, "Syndication networks and the spatial distribution of venture capital investment," *The American Journal of Sociology*, vol. 106, no. 6, pp. 1546–1588, May 2001.

[4] B. Kogut, P. Urso, and G. Walker, "Emergent properties of a new financial market: American venture capital syndication, 1960-2005," *Management Science*, vol. 53, no. 7, pp. 1181–1198, July 2007.

[5] W. Powell, D. White, K. Koput, and J. Owen-Smith, "Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences," *American Journal of Sociology*, vol. 110, no. 4, pp. 1132–1205, January 2005.

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