# Learning Triadic Influence in Large Social Networks

**Chenhui Zhang[1], Sida Gao[1], Jie Tang[1], Tracy Xiao Liu[1], Zhanpeng Fang[1], Xu Cheng[2]**

[1]Tsinghua University; [2]Tencent Corporation

## Triadic Influence: What role do the triadic structures among one's friends play in influencing his/her behavior?

- Social influence occurs when one's opinions, emotions, or behaviors are affected by others. *Triadic Influence Analysis* focuses on the influence from the perspective of structure, particularly on triads.
- Triads are the simplest group structure in networks as well as the cornerstone for study network formation. There are different triadic influence patterns when the label information and the strength of relationships are available.
- Experiment results show that by leveraging the triadic influence, the prediction accuracy can be significantly improved on *retweeting* behavior in Weibo (+13.84%) and *paying* behavior in CrossFire (+1.47%).

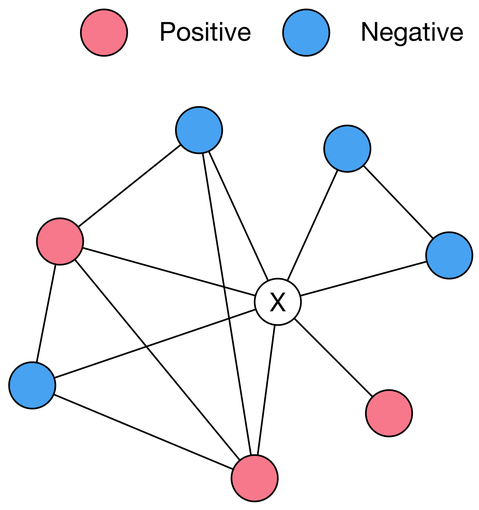
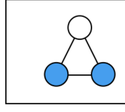
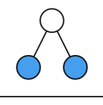



Table 1 Number of different triadic structures in X's ego network

| Closed Triadic Structures | | Open Triadic Structures | |
|---|---|---|---|
|  | 1 |  | 2 |
|  | 4 |  | 8 |
|  | 1 |  | 6 |

## Empirical analysis on different triadic influence patterns

**Dataset:**

***Weibo***---The data set comprises the following-relationship network, user profiles, and microblog diffusion episodes.

***CrossFire***---The data set comprises user profiles and all kinds of logs such as login, buying item and playing game.

**OLS Analysis:**

The variables in linear regression model include

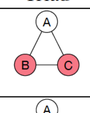
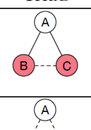
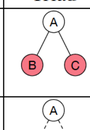
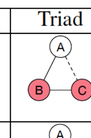
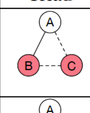
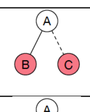
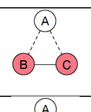
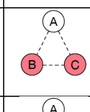
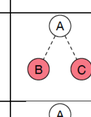
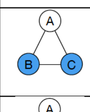
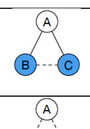
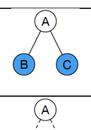
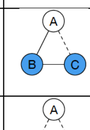
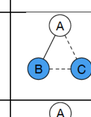
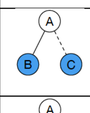
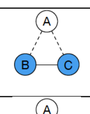
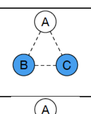
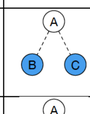
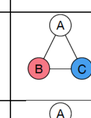
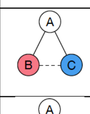
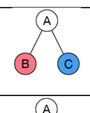
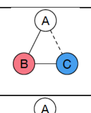
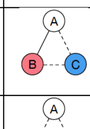
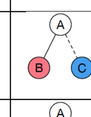
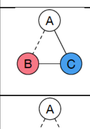
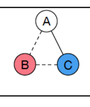
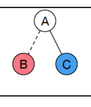
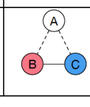
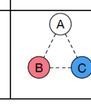
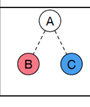
- Basic Features---Extracted from user profiles.
- Neighborhood Features---Number of positive (or negative) friends with strong (or weak) relationships.
- Triadic Features---Number of 30 triadic influence patterns according to the label information and the strength of relationships

**Regression coefficients on Weibo dataset:**

REGRESSION ANALYSIS FOR NEIGHBORHOOD FEATURES

| | Edge | B+E | B+E+Triads |
|---|---|---|---|
| #PosStrong | 0.0840*** (0.002) | 0.0544*** (0.002) | 0.0602*** (0.006) |
| #PosWeak | 0.1268*** (0.003) | 0.0700*** (0.002) | 0.0679*** (0.012) |
| #NegStrong | 0.0355*** (0.001) | 0.1670*** (0.002) | 0.0490*** (0.009) |
| #NegWeak | -0.0490*** (0.001) | -0.0040** (0.002) | 0.0201 (0.023) |
| $R^2$ | 0.041 | 0.242 | 0.301 |

REGRESSION ANALYSIS FOR 30 KINDS OF TRIADS

| No. | Triad | Coef | No. | Triad | Coef | No. | Triad | Coef | No. | Triad | Coef | No. | Triad | Coef |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 |  | 0.0827*** (0.003) | 2 |  | 0.0110*** (0.004) | 3 |  | -0.0543*** (0.003) | 4 |  | 0.0004 (0.004) | 5 |  | 0.0429*** (0.003) |
| 6 |  | -0.0587*** (0.003) | 7 |  | 0.0205*** (0.002) | 8 |  | 0.0313*** (0.002) | 9 |  | -0.0283*** (0.003) | 10 |  | 0.0168*** (0.002) |
| 11 |  | -0.0091*** (0.002) | 12 |  | 0.0861*** (0.004) | 13 |  | -0.0563*** (0.001) | 14 |  | -0.0221*** (0.002) | 15 |  | 0.0157*** (0.003) |
| 16 |  | 0.0167*** (0.003) | 17 |  | 0.0164*** (0.003) | 18 |  | -0.0184* (0.010) | 19 |  | -0.0263*** (0.002) | 20 |  | 0.0066*** (0.002) |
| 21 |  | -0.0001 (0.002) | 22 |  | 0.0099*** (0.002) | 23 |  | 0.0534*** (0.002) | 24 |  | -0.0054** (0.002) | 25 |  | 0.0089*** (0.001) |
| 26 |  | -0.0783*** (0.002) | 27 |  | 0.0818*** (0.002) | 28 |  | 0.0130*** (0.002) | 29 |  | 0.0494*** (0.002) | 30 |  | -0.0772*** (0.003) |

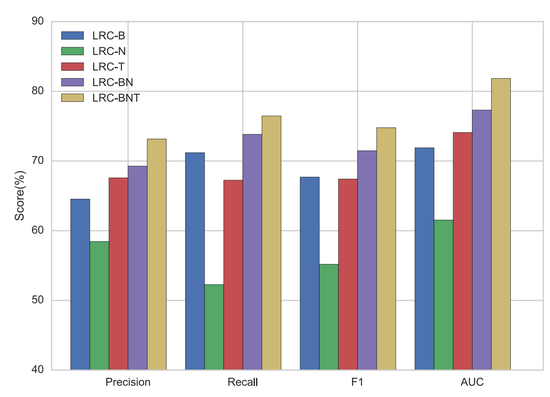
- Because of the Triadic features, the #NegWeak edges become insignificant.
- Although Neighborhood features are included in the model, most of the triadic influence patterns are significant.
- Most of the open triads result in negative coefficients even when the neighbors are positive (triad 3, 6, 9), which is quite different from closed triads.
- The strength of relationship plays an important role in influencing one's behavior (Triad 1 vs. Triad 7), but it is not necessarily positively correlated.
- The relationship between B and C also have influence on the retweeting behavior of A when compared the (Triad 1 vs. Triad 2).

## Experiment results on behavior prediction

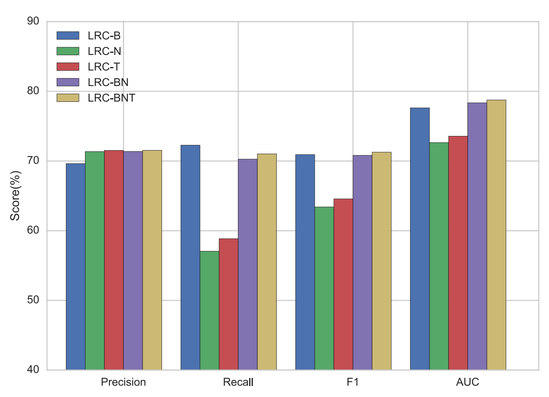
PERFORMANCE OF BEHAVIOR PREDICTION ON TWO DATASETS. (%)

| Dataset | Model | Precision | Recall | F1 | AUC |
|---|---|---|---|---|---|
| Weibo | LRC-B | 64.54 | 71.19 | 67.70 | 71.90 |
| | LRC-N | 58.46 | 52.27 | 55.19 | 61.53 |
| | LRC-T | 67.59 | 67.25 | 67.42 | 74.09 |
| | LRC-BN | 69.27 | 73.83 | 71.47 | 77.31 |
| | LRC-BNT | **73.16** | **76.46** | **74.78** | **81.85** |
| CrossFire | LRC-B | 69.63 | **72.27** | 70.92 | 77.62 |
| | LRC-N | 71.35 | 57.06 | 63.41 | 72.64 |
| | LRC-T | 71.52 | 58.85 | 64.57 | 73.56 |
| | LRC-BN | 71.37 | 70.27 | 70.81 | 78.35 |
| | LRC-BNT | **71.53** | 71.02 | **71.27** | **78.76** |

Retweeting behavior prediction in Weibo




Retweeting buying prediction in Crossfire




- LRC-B/LRC-N/LRC-T: Logistic regression classifier with only Basic/Neighborhood/Triadic features
- LRC-BN: Logistic regression classifier with Basic and Neighborhood features
- LRC-BNT: Logistic regression classifier with all of the three features

- Although Neighborhood features are much less powerful when used alone, it can improve the performance of Basic features significantly when they are combined together.
- Triadic features alone is almost as good as combining Basic and Neighborhood features.
- By adding triadic features to traditional methods, the predictive power of our model increases significantly on both datasets.