

Balance in Signed Bipartite Networks

Cheolhee Jung
DMLAB

Tyler Derr, Cassidy Johnson, Yi Chang, Jiliang Tang
"Balance in Signed Bipartite Networks"

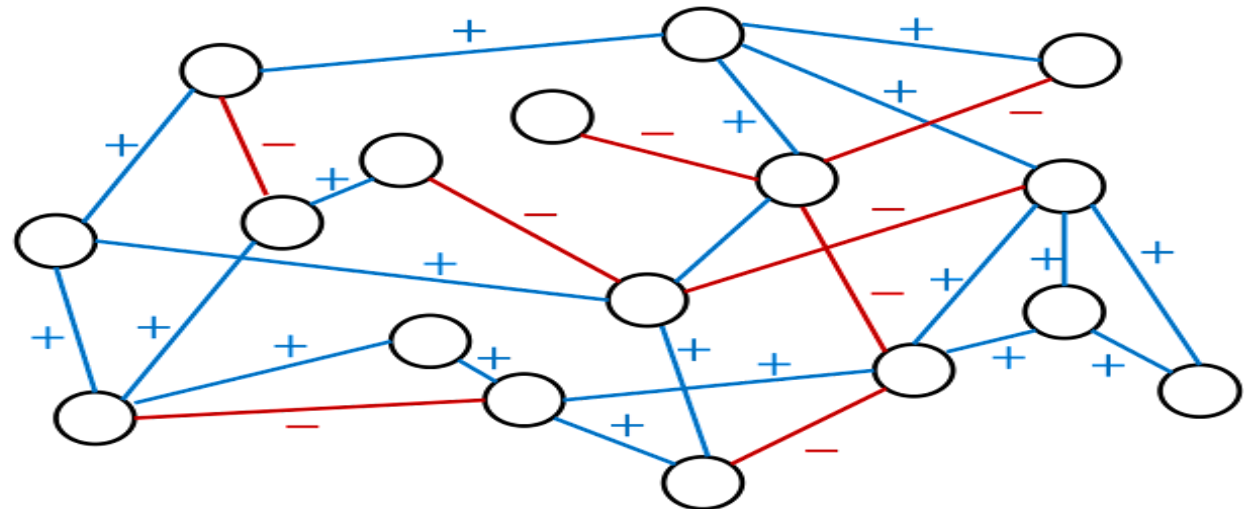


Contents

- **Background**
- **Problem definition**
- **Main ideas**
- **Proposed method**
- **Experiment**
- **Conclusion**
 - **2 strong points**
 - **2 weak points**
 - **Improve or resolve those weak point**

Background

- ▷ 데이터가 네트워크 형태로 표현
- ▷ 소셜 미디어와 전자 상거래의 확산으로 긍정적 및 부정적 관계를 포함한 부호화 네트워크의 중요성 부각
- ▷ 기존 네트워크 분석 방법은 이러한 복잡성을 다루기에 부족하여, 새로운 이론과 알고리즘이 필요



Problem definition

Q. 기존 분석 방식으로는 한계가 있는
부호화 이분 네트워크를 어떻게 분석할까?

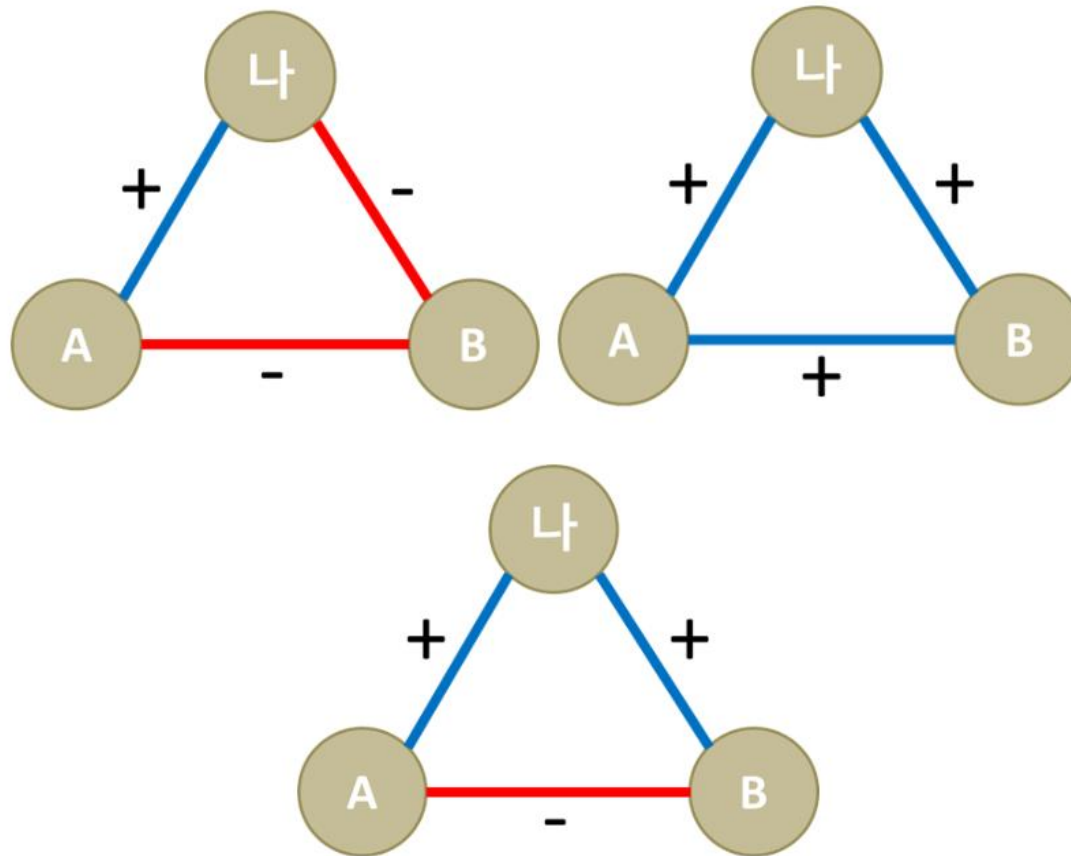


Main ideas

- ▷ 균형 이론과 부호화 나비 네트워크를 통한 새로운 분석 방법
- ▷ 균형 이론을 부호화 이분 네트워크 분석에 적용하여, 네트워크 내에서 긍정적 및 부정적 연결의 패턴을 이해하고 예측하는 새로운 방법론을 개발

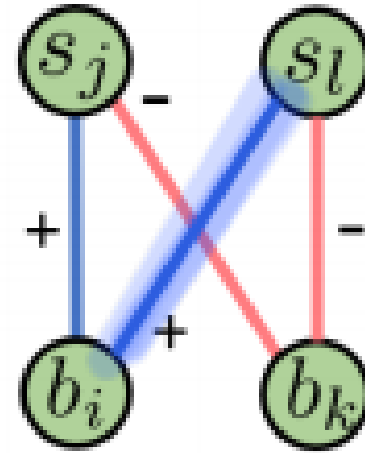
Main ideas

balance theory



Main ideas

butterflies



부호화 이분 네트워크에서의 가장 작은 사이클
 ➡ 균형이론을 적용

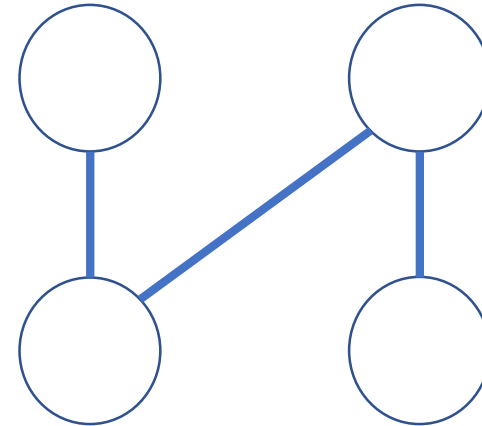
Proposed method

- ▷ 부호화 애벌레 기반 분류기
- ▷ 저순위 행렬 분해
- ▷ 랜덤 워크 기반 부호 예측

위 3가지 방법은 부호화 이분 네트워크에서 누락된 링크의 부호를 예측하기 위해 부호화 나비를 통해 균형 이론을 활용

Proposed method

부호화 애벌레 기반 분류기



하나의 링크가 누락된 네트워크

부정적 연결 짝 수 : 균형 잡힌 경로

부정적 연결 홀 수 : 불균형한 경로

Proposed method

부호화 애벌레 기반 분류기

▷ 특성 추출

1. 노드 차수 기반 특성 x_{ij}^d
 - 구매자와 판매자 개인의 긍정적 및 부정적 차수 정보 포함
2. 부호화 애벌레 기반 특성 x_{ij}^{sc}
 - 구매자와 판매자를 연결하는 8가지 가능한 부호화 애벌레 유형의 수를 특성으로 사용

▷ 분류기 학습 및 예측

- 추출된 특성을 기반으로, 로지스틱 회귀와 같은 분류 알고리즘을 사용하여 모델을 학습
- 학습된 모델을 이용하여 알려지지 않은 링크의 부호를 예측

Proposed method

저차원 행렬 분해

▷ 기본 개념 및 최적화 문제

목적: 부호화 이분 네트워크에서의 링크 부호 예측

모델: 이분 인접 행렬 B 를 사용하여 구매자와 판매자 간의 잠재 표현을 찾는
저차원 행렬 분해

최적화 문제:
$$\min_{U, V} \sum_{(b_i, s_j) \in \mathcal{E}} \max \left(0, 1 - \mathbf{B}_{ij}(\mathbf{u}_i^\top \mathbf{v}_j) \right)^2 + \lambda \left(|\mathbf{U}|_F^2 + |\mathbf{V}|_F^2 \right)$$

$\mathcal{E} = \{(b_i, s_j) | \mathbf{B} \neq 0\}$: 실제 연결 쌍

$\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{n_B}] \in \mathbb{R}^{d \times n_B}$ and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{n_S}] \in \mathbb{R}^{d \times n_S}$: 구매자와 판매자 잠재 행렬

\mathbf{B}_{ij} : 실제 부호 $(\mathbf{u}_i^\top \mathbf{v}_j)$: 구매자 i , 판매자 j 사이의 관계의 강도와 부호를 예측

Proposed method

저차원 행렬 분해

▷ 균형 이론 기반 확장 모델(행렬 분해만으로는 균형 이론을 따르는 예측 보장 못함)

확장된 목적 함수 :

$$\min_{\mathbf{U}, \mathbf{V}} \sum_{(b_i, s_j) \in \mathcal{E}} \max \left(0, 1 - \mathbf{B}_{ij}(\mathbf{u}_i^\top \mathbf{v}_j) \right)^2 + \lambda \left(|\mathbf{U}|_F^2 + |\mathbf{V}|_F^2 \right) \\ + \alpha \sum_{(b_i, s_j) \in \hat{\mathcal{E}}_i^+} \max \left(0, 1 - \hat{\mathbf{S}}_{ij}(\mathbf{u}_i^\top \mathbf{v}_j) \right)^2 \\ + \beta \sum_{(b_i, s_j) \in \hat{\mathcal{E}}_i^-} \max \left(0, 1 - \hat{\mathbf{S}}_{ij}(\mathbf{u}_i^\top \mathbf{v}_j) \right)^2$$

$\hat{\mathcal{E}}_i^+ \quad \hat{\mathcal{E}}_i^-$: 균형 이론에 의해 제안된 양성, 음성 링크 집합

$\alpha \quad \beta$: 양성 및 음성 링크의 중요도 조절

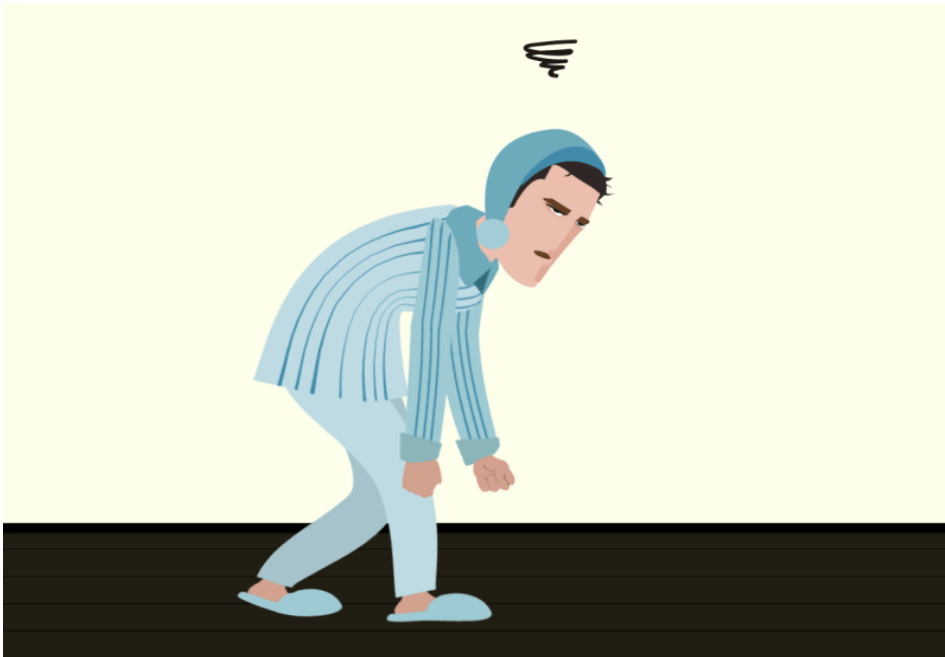
$\hat{\mathbf{S}}_{ij}$: 균형 이론에 기반하여 제안된 링크 부호

Proposed method

랜덤 워크 기반 부호 예측

Lazy random walk : 랜덤 워커가 다음 스텝에서 현재 노드에 머물 확률을 포함

➡ 부호화 이분 네트워크 정상 분포 미도달 문제 해결



Proposed method

랜덤 워크 기반 부호 예측

▷ 기본 개념

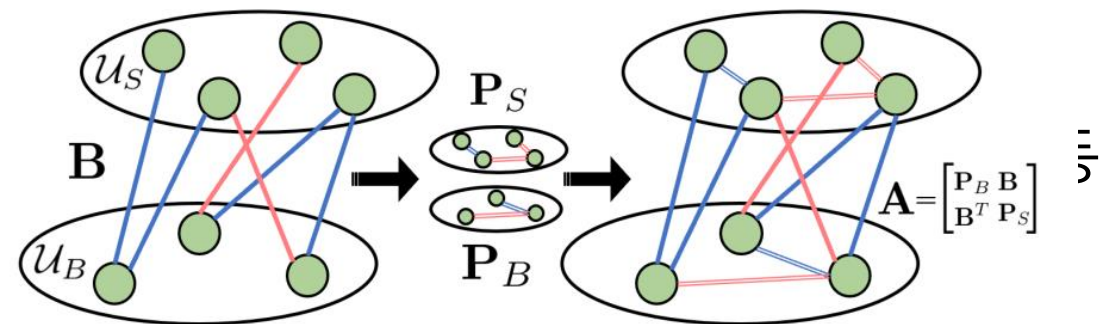
목적: 서명된 이분 네트워크에서 미지의 링크 부호를 예측

핵심 요소:

한 모드 투영 네트워크: 구매자 또는 판매자 기반으로 생성된 네트워크,
공통된 연결을 기반으로 노드 간의 숨겨진 관계를 찾음

균형 이론 적용: 네트워크 내 긍정적 및 부정적 연결의 균형을 고려하여
안정성과 사회적 선

랜덤 워크 : 특정 노드에서 시작하여
- **특징 :** 재시작 확률로 인해 시킨



정상 분포 : 랜덤 워크에서 장기적으로 시스템이 수렴하는 확률 분포

Proposed method

랜덤 워크 기반 부호 예측

▷ 한 모드 투영 네트워크 구성

목적 : 같은 유형의 노드 간 암시적 연결 생성

투영 네트워크 : 구매자 간(PB), 판매자 간(PS) 인접 행렬을 통해 표현

구성 방식 : 공통 판매자/구매자를 기반으로 구매자/판매자 간 연결

$$P_{Bij} = P_{Bji} = ns_{ij}^A - ns_{ij}^D$$

$$P_{Bij} = \begin{cases} 0 & \delta_n < ns_{ij}^A - ns_{ij}^D < \delta_p \\ ns_{ij}^A - ns_{ij}^D & \text{otherwise} \end{cases}$$

ns_{ij}^A : 동의하는 공통 판매자의 수

ns_{ij}^D : 불일치하는 판매자의 수

Proposed method

랜덤 워크 기반 부호 예측

▷ 랜덤 워크 수행 및 부호 예측

통합된 인접 행렬 A 구성 : $A = \begin{bmatrix} \hat{P}_B & \omega \hat{B} \\ \omega \hat{B}^T & \hat{P}_S \end{bmatrix}$

\hat{P}_B \hat{P}_S : 행 정규화된 투영 인접 행렬

\hat{B} : 행 정규화된 이분 인접 행렬

ω : 실제 연결(B), 간접적 연결(투영 인접) 사이 중요도

행 정규화 : 네트워크 연결구조를 표준화하고 정규화 하는 단계

➡ 각 노드의 연결 강도를 비교 가능한 기준으로 조정 가능

Proposed method

랜덤 워크 기반 부호 예측

▷ 부호 예측

부호 예측 행렬 Y 를 업데이트 $Y_{ij} = \sum_k \hat{A}_{ik} Y_{kj}$

Y_{ij} : 노드 i 와 j 사이의 부호 예측값

\hat{A}_{ik} : 노드 간의 이동 확률

균형 이론을 기반으로 부호 예측

$\hat{A}_{ik} Y_{kj} > 0$ Y_{ij} 를 증가 시켜 긍정 관계

$\hat{A}_{ik} Y_{kj} < 0$ Y_{ij} 를 감소 시켜 부정 관계

Proposed method

랜덤 워크 기반 부호 예측

▷ 재시작 확률을 포함한 폐쇄 형태의 해

$$Y = (1 - c)(I - c\hat{A})^{-1}$$

수치 해석

폐쇄 형태의 해 $(I - c\hat{A})^{-1}$:

랜덤 워크가 각 단계에서 재시작할 가능성을 고려하여, 네트워크의 모든 노드 간의 연결 가능성을 계산

계수 $(1 - c)$:

재시작하지 않고 계속해서 랜덤 워크를 수행할 확률

c 가 0에 가까울수록 이 값은 1에 가까워지며, 랜덤 워크가 네트워크 내에서 더 넓은 범위를 탐색

Experiment

Table 4: Link Sign Prediction Results in terms of (AUC,F1).

Sign Prediction Method	Bonanza	U.S. Senate	U.S. House
SCd	(0.553 , 0.959)	(0.638 , 0.654)	(0.625 , 0.635)
SCsc	(0.664 , 0.674)	(0.812 , 0.823)	(0.827 , 0.837)
MF	(0.593 , 0.903)	(0.792 , 0.812)	(0.831 , 0.846)
MFwBT	(0.608 , 0.905)	(0.814 , 0.827)	(0.834 , 0.848)
LazyRW	(0.547 , 0.979)	(0.808 , 0.821)	(0.815 , 0.827)
SBRW	(0.582 , 0.949)	(0.836 , 0.849)	(0.846 , 0.858)

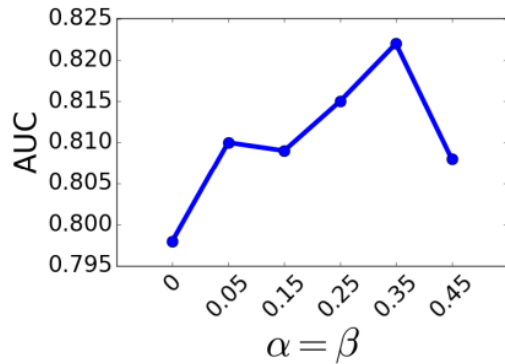
실험 개요

- 부호화 이분 네트워크에서의 부호 예측 성능을 AUC 및 F1 점수로 평가
- 비교된 방법: SCd, SCsc, MF, MFwBT, LazyRW, SBRW
- 데이터셋: Bonanza, U.S. Senate, U.S. House

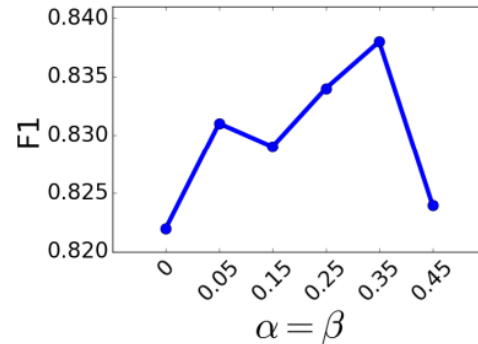
주요 발견점

- 모든 데이터셋에서 동일한 방법이 가장 우수한 성능을 보이지 않음
- 균형 이론에 기반한 방법들(SCsc, MFwBT, SBRW)이 기존 방법들(SCd, MF, LazyRW)보다 AUC에서는 일관되게 우수하지만, F1 점수는 경우에 따라 다름

Experiment

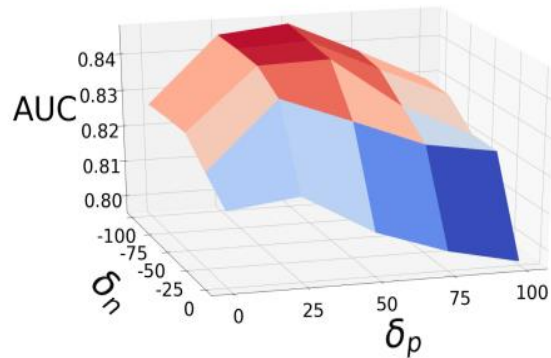


(a) MFwBT (AUC)

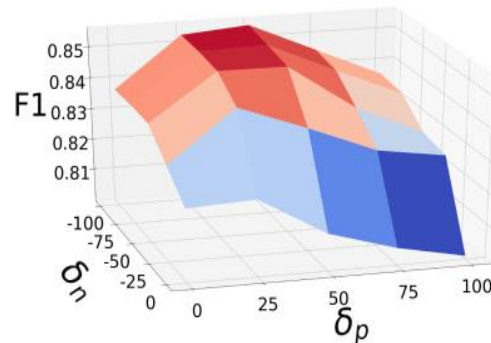


(b) MFwBT (F1)

Figure 4: Parameter Sensitivity on α and β in MFwBT on the U.S. Senate dataset.



(a) SBRW (AUC)



(b) SBRW (F1)

Figure 5: Parameter Sensitivity on δ_p and δ_n in SBRW on the U.S. House dataset.

MFwBT 방법

개요

- MFwBT는 부호 예측을 위해 균형 이론과 저차원 행렬 분해 기법을 결합한 방법
- α 와 β 는 부호화 나비를 기반으로 추출한 긍정적 및 부정적 링크의 훈련 기여도를 조절

파라미터 민감도

- 균형 이론을 고려하지 않는 경우($\alpha = \beta = 0$)보다 부호 예측 성능이 향상되지만, 너무 높은 값을 사용하는 것은 지양

SBRW 방법

개요

- SBRW는 부호화 이분 네트워크에서 랜덤 워크를 수행하여 링크 부호를 예측하는 방법

파라미터 민감도

- 낮은 지지도를 가진 암시적 긍정적 링크는 포함하는 것이 유용하지만, 부정적 링크는 균형 이론 지지가 낮은 경우 제외하는 것이 성능 향상에 도움

Conclusion

2 strong point

1. 복잡한 네트워크를 분석하는 새로운 관점을 제공
2. 부호화 나비를 효과적으로 활용하여 링크 부호를 예측하는데 있어 우수한 성능을 보임

Conclusion

2 weak point

1. 네트워크의 시간적 변화를 고려하지 않음
2. 노드 및 링크의 유형이 한정 돼 있음

Conclusion

Improve or resolve those weak point

동적 네트워크 분석 적용

- 시간 정보를 포함한 동적 네트워크 분석 방법을 개발

다양한 네트워크 유형에 대한 연구 확장

- 다양한 노드 유형과 링크 패턴을 포함하는
네트워크를 연구하여, 기존 방법론의 적용 범위를 확장



Thank you!