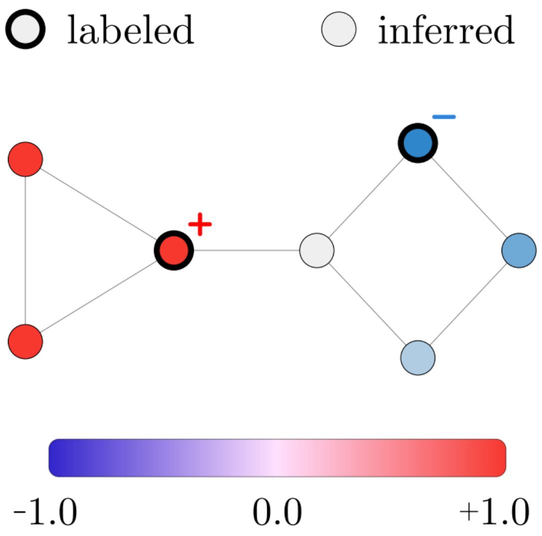
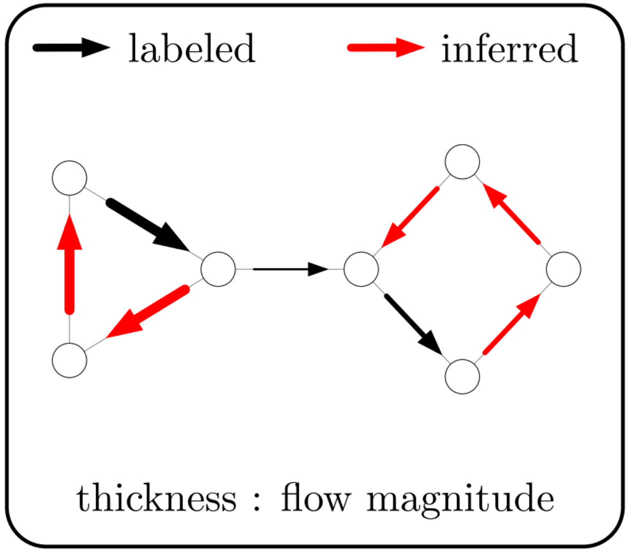




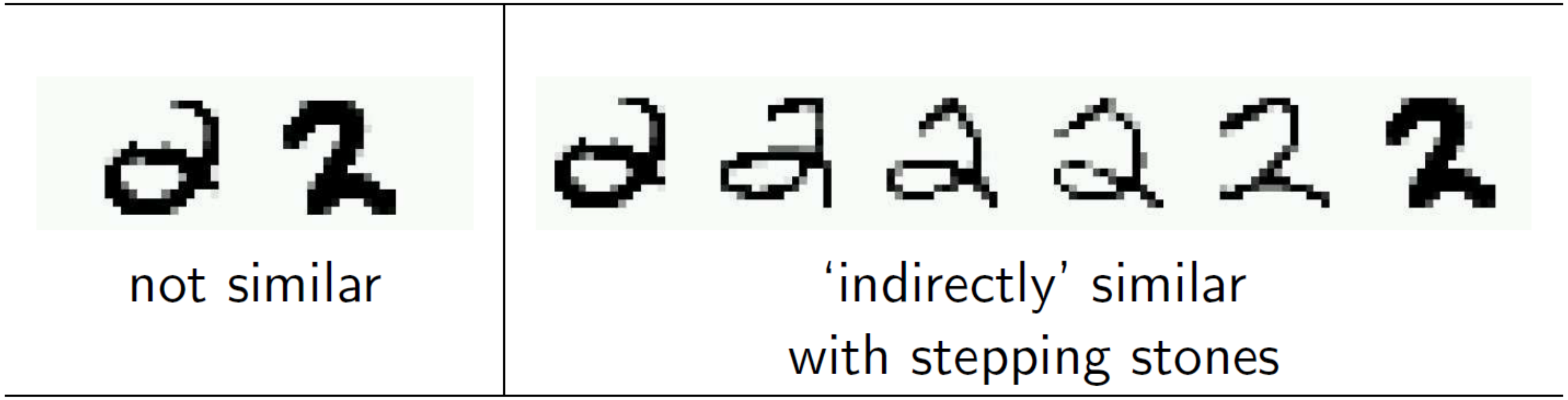
Semi-Supervised Learning: Graph-based SSL

Pilsung Kang
School of Industrial Management Engineering
Korea University

Graph-based SSL

Zhu (2007)

- Example: Handwritten digit recognition with pixel-wise Euclidean distance



- Assumption
  - ✓ A graph is given on the labeled and unlabeled data
  - ✓ Instances connected by heavy edge tend to have the same label

Graph-based SSL

Zhu (2007)

- Graph construction
  - ✓ Nodes: $\mathbf{X}_l \cup \mathbf{X}_u$
  - ✓ Edges: similarity weights computed from features, e.g.,
    - k-nearest-neighbor graph, unweighted (0, 1 weights)
    - fully connected graph, weight decays with distance

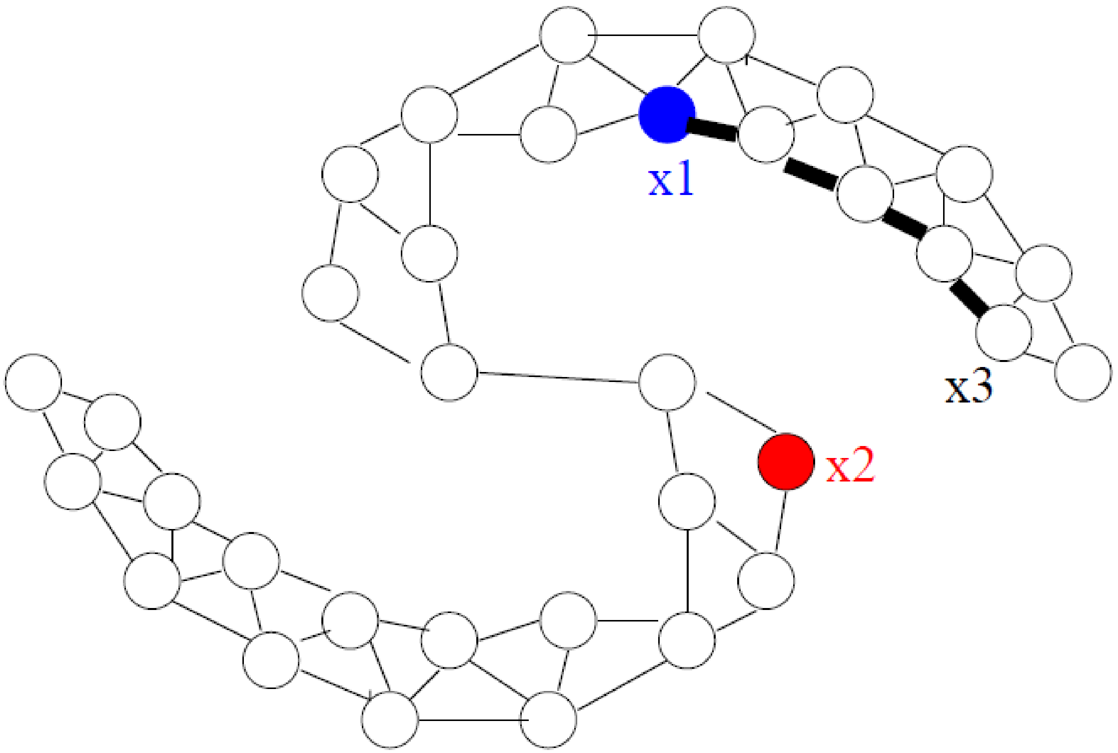
$$w_{ij} = \exp\left(\frac{-||\mathbf{x}_i - \mathbf{x}_j||^2}{\sigma^2}\right)$$

    - ε-radius graph
  - ✓ Assumption: Instances connected by heavy edge tend to have the same label

Graph-based SSL

Zhu (2007)

- Assumption: Instances connected by heavy edge tend to have the same label



Graph-based SSL

Zhu (2007)

- The mincut algorithm
  - ✓ Fix $\mathbf{y}_l$, find $\mathbf{y}_u \in \{0,1\}^{n-l}$ to minimize $\sum_{i,j} w_{ij}|y_i - y_j|$
  - ✓ Equivalently, solve the optimization problem

$$\min_{\mathbf{y}\in\{0,1\}^n} \infty \sum_{i=1}^{l} (y_i - \mathbf{y}_{li})^2 + \sum_{i,j} w_{ij}(y_i - y_j)^2$$

  - ✓ Combinatorial problem, but has polynomial time solution

고려대학교 KOREA UNIVERSITY

DSBA Data Science & Business Analytics

Graph-based SSL

Zhu (2007)

- Harmonic function
  - ✓ Relaxing discrete labels to continuous values in $\mathbb{R}$, the harmonic function f satisfies

$$f(\mathbf{x}_i) = y_i \text{ for } i = 1, ..., l$$

  - f minimizes the energy

$$\sum_{i \sim j} w_{ij} (f(\mathbf{x}_i) - f(\mathbf{x}_j))^2$$

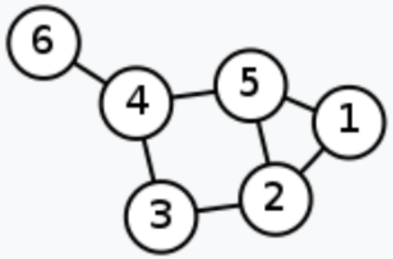
  - the mean of a Gaussian random field
  - average of neighbors

$$f(\mathbf{x}_i) = \frac{\sum_{j \sim i} w_{ij} f(\mathbf{x}_j)}{\sum_{j \sim i} w_{ij}}, \ \forall \ \mathbf{x}_i \in \mathbf{X}_u$$

Graph-based SSL

Zhu (2007)

- The graph Laplaican
  - ✓ We can also compute f in closed form using the graph Laplacian
    - $n \times n$ weight matrix $\mathbf{W}$ on $\mathbf{X}_l \cup \mathbf{X}_u$
      - Symmetric, non-negative
    - Diagonal degree matrix: $$\mathbf{D} : \mathbf{D}_{ii} = \sum_{j=1}^{n} \mathbf{W}_{ij}$$
    - Graph Laplacian matrix $$\mathbf{\Delta} = \mathbf{D} - \mathbf{W}$$

Labeled graph	Degree matrix	Adjacency matrix	Laplacian matrix
	$\begin{pmatrix} 2&0&0&0&0&0\\ 0&3&0&0&0&0\\ 0&0&2&0&0&0\\ 0&0&0&3&0&0\\ 0&0&0&0&3&0\\ 0&0&0&0&0&1 \end{pmatrix}$	$\begin{pmatrix} 0&1&0&0&1&0\\ 1&0&1&0&1&0\\ 0&1&0&1&0&0\\ 0&0&1&0&1&1\\ 1&1&0&1&0&0\\ 0&0&0&1&0&0 \end{pmatrix}$	$\begin{pmatrix} 2&-1&0&0&-1&0\\ -1&3&-1&0&-1&0\\ 0&-1&2&-1&0&0\\ 0&0&-1&3&-1&-1\\ -1&-1&0&-1&3&0\\ 0&0&0&-1&0&1 \end{pmatrix}$

Graph-based SSL

Zhu (2007)

- The graph Laplaican
  - ✓ We can also compute f in closed form using the graph Laplacian
    - $n \times n$ weight matrix $\mathbf{W}$ on $\mathbf{X}_l \cup \mathbf{X}_u$
      - Symmetric, non-negative
    - Diagonal degree matrix: $\mathbf{D} : \mathbf{D}_{ii} = \sum_{j=1}^{n} \mathbf{W}_{ij}$
    - Graph Laplacian matrix $\boldsymbol{\Delta} = \mathbf{D} - \mathbf{W}$
    - The energy can be rewritten as

$$\frac{1}{2} \sum_{i \sim j} w_{ij} (f(\mathbf{x}_i) - f(\mathbf{x}_j))^2 = \mathbf{f}^T \boldsymbol{\Delta} \mathbf{f}$$

Graph-based SSL

Zhu (2007)

- Harmonic solution with Laplacian
  - ✓ The harmonic solution minimizes energy subject to the given labels

$$\min_{\mathbf{f}} \infty \sum_{i=1}^{l} (f(\mathbf{x}_i) - y_i)^2 + \mathbf{f}^T \boldsymbol{\Delta} \mathbf{f}$$

  - Partition the Laplacian matrix

$$\boldsymbol{\Delta} = \begin{bmatrix} \boldsymbol{\Delta}_{ll} & \boldsymbol{\Delta}_{lu} \\ \boldsymbol{\Delta}_{ul} & \boldsymbol{\Delta}_{uu} \end{bmatrix}$$

  - Harmonic solution

$$\mathbf{f}_u = -\boldsymbol{\Delta}_{uu}^{-1} \boldsymbol{\Delta}_{ul} \mathbf{y}_l$$

  - The normalized Laplacian is often used

$$\mathcal{L} = \mathbf{D}^{-\frac{1}{2}} \boldsymbol{\Delta} \mathbf{D}^{-\frac{1}{2}} = \mathbf{I} - \mathbf{D}^{-\frac{1}{2}} \mathbf{W} \mathbf{D}^{-\frac{1}{2}}$$

Graph-based SSL

Zhu (2007)

- Problems with harmonic solution
  - ✓ It fixes the given labels $\mathbf{y}_l$
    - What if some labels are wrong?
    - Want to be flexible and disagree with given labels occasionally
  - ✓ It cannot handle new test points directly
    - f is only defined on $\mathbf{X}_u$
    - We have to add new test points to the graph, and find a new harmonic solution

➢ Allow $f(\mathbf{X}_l)$ to be different from $\mathbf{y}_l$ but penalize it

➢ Introduce a balance between labeled data fit and graph energy

$$\min_{\mathbf{f}} \sum_{i=1}^{l} (f(\mathbf{x}_i) - y_i)^2 + \lambda \mathbf{f}^T \boldsymbol{\Delta} \mathbf{f}$$

Graph-based SSL

- Solution

$$\min(\mathbf{f}-\mathbf{y})^T(\mathbf{f}-\mathbf{y})+\lambda\mathbf{f}^T\mathbf{\Delta}\mathbf{f} \qquad \mathbf{y}=[\mathbf{y}_l; \underbrace{0; \quad \ldots \quad ; \quad 0}_{N.\text{of unlabeled examples}} \quad]$$

$$\frac{\partial E}{\partial \mathbf{f}}=(\mathbf{f}-\mathbf{y})+\lambda\mathbf{\Delta}\mathbf{f}=0$$

$$(\mathbf{I}+\lambda\mathbf{\Delta})\mathbf{f}=\mathbf{y} \quad \Rightarrow \quad \mathbf{f}=(\mathbf{I}+\lambda\mathbf{\Delta})^{-1}\mathbf{y}$$

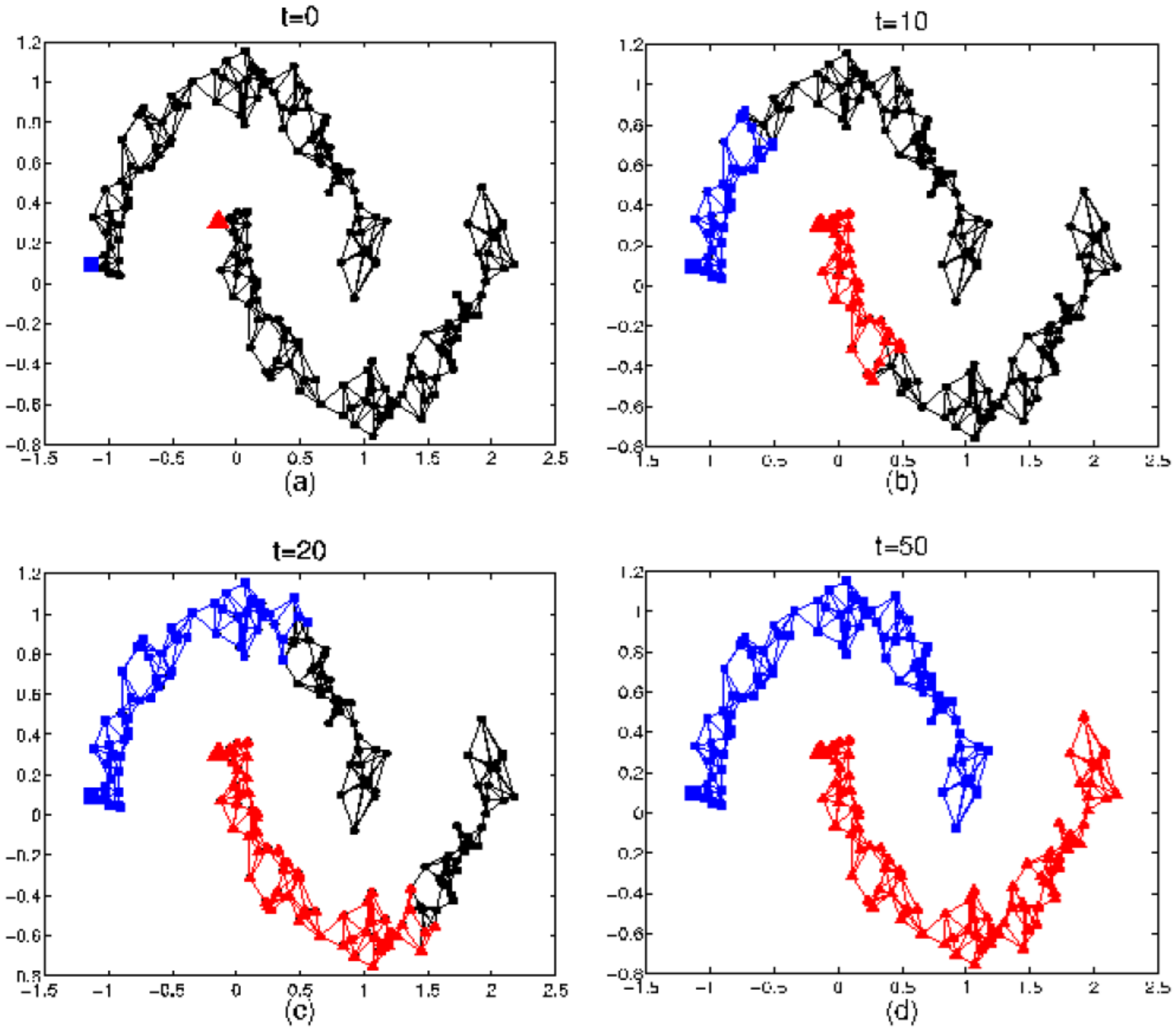
✓ If λ is large, then the effect of $\lambda\mathbf{\Delta}$ increases → more focused on the smoothness

✓ If λ is small, then the effect of $\lambda\mathbf{\Delta}$ decreases → more focused on the accuracy of labeled data

Graph-based SSL

Zien (2008)

- Examples: Label Propagation

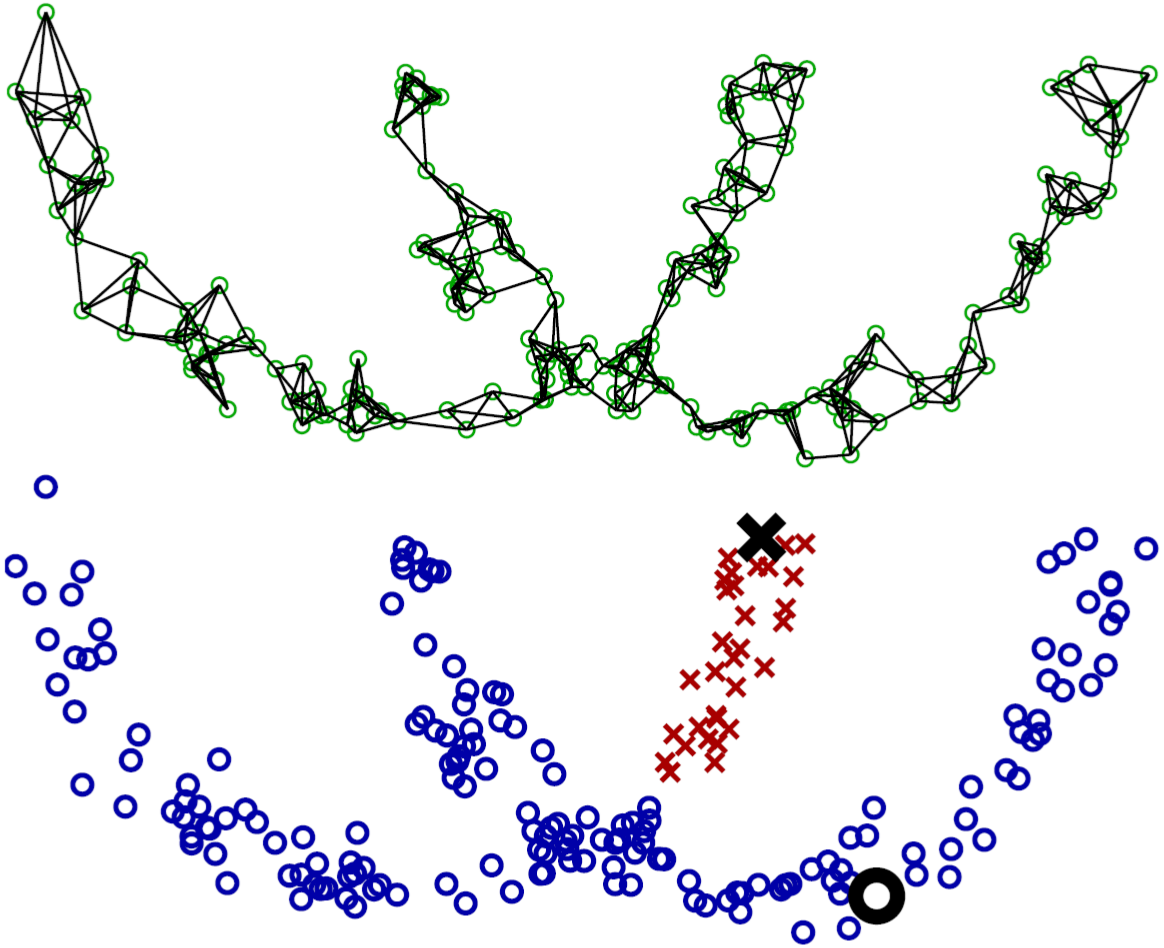


Graph-based SSL

Zhu (2009)

- When the graph assumption is wrong

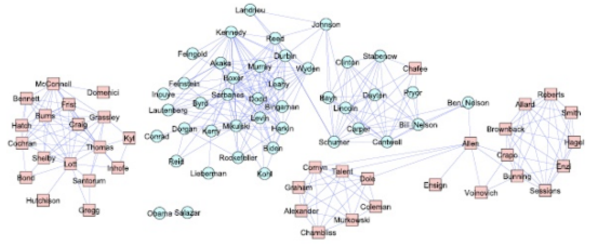
"colliding two moons"



Graph-based SSL: Word Sentiment Propagation

- Framework

❖ Framework



Text + scaled score
데이터 수집

02

임베딩 공간에서 거리기반
네트워크 구축

04

추정한 감성 점수
성능 평가

데이터 준비 | Word2Vec | Construct network | Graph based SSL | 성능 평가

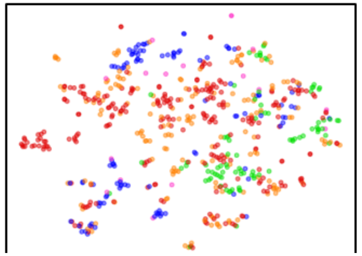
01

고차원으로 표현된 단어를
저차원으로 임베딩

03

명백한 label을 미리 선정/할당
label propagation을 이용
score 추정

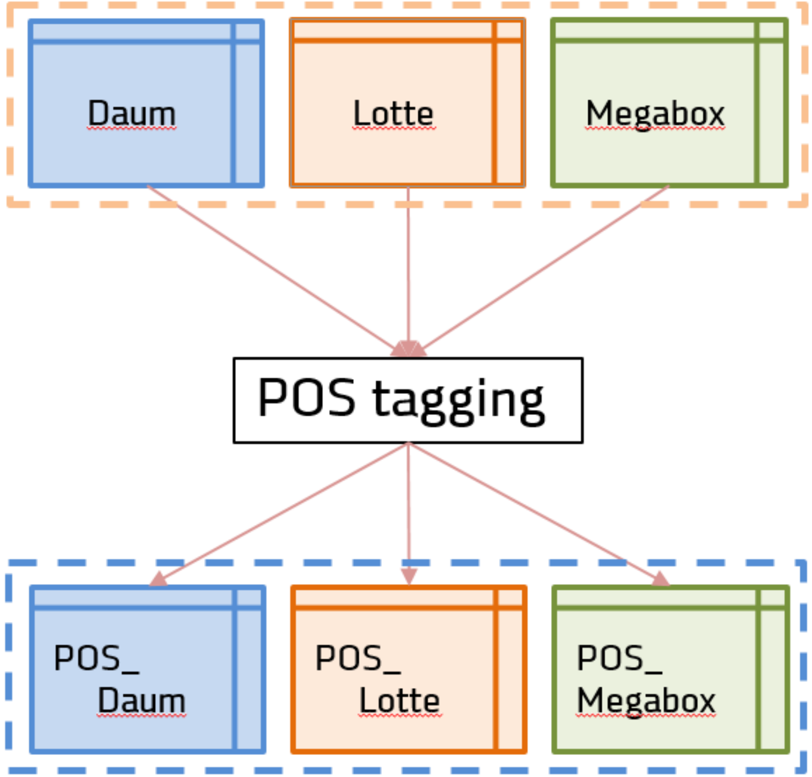
05



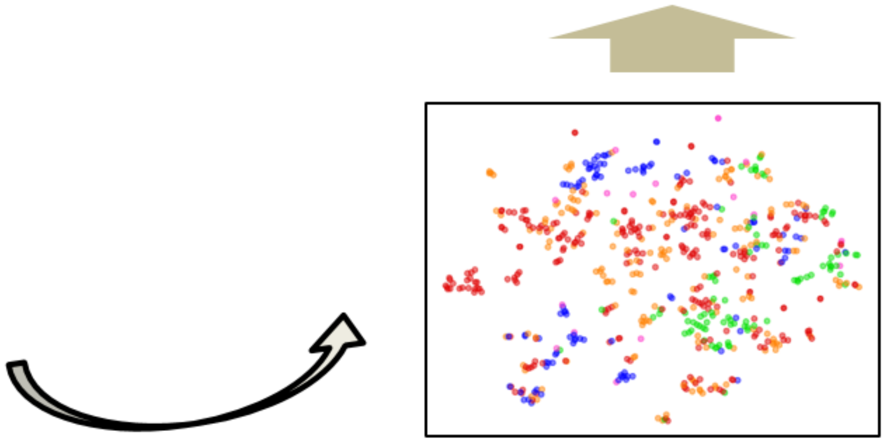
Term	Score
재미있다	1
…	…
지루하다	-1

Graph-based SSL: Word Sentiment Propagation

- Framework



Term	d_1	d_2	…	d_n
명작	0.234	0.891	…	0.963
아쉽다	0.477	0.443	…	0.564
…	…	…	…	…
연기력	0.119	0.863	…	0.787



고려대학교 KOREA UNIVERSITY

DSBA Data Science & Business Analytics

Graph-based SSL: Word Sentiment Propagation

- Framework

데이터 준비 → Word2Vec → Construct network → Graph based SSL → 성능 평가

Term	d_1	d_2	…	d_n
명작	0.234	0.891	…	0.963
아쉽다	0.477	0.443	…	0.564
…	…	…	…	…
연기력	0.119	0.863	…	0.787

Distance Matrix

중심성

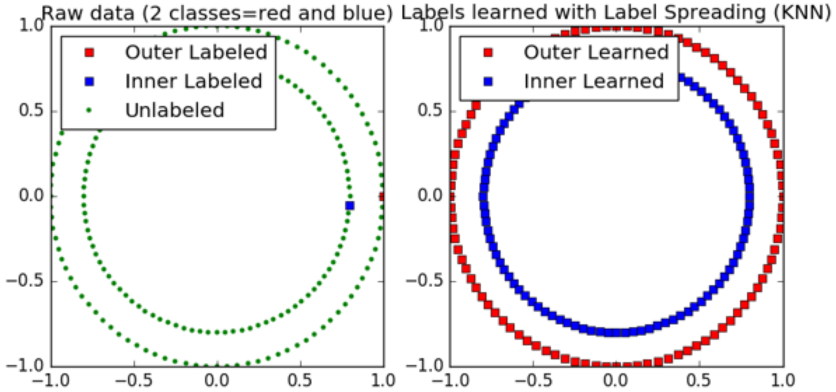


그림 출처 : scikit learn

Pre-labeled word

상위 50	상위 50
긍정 (331)	부정 (219)

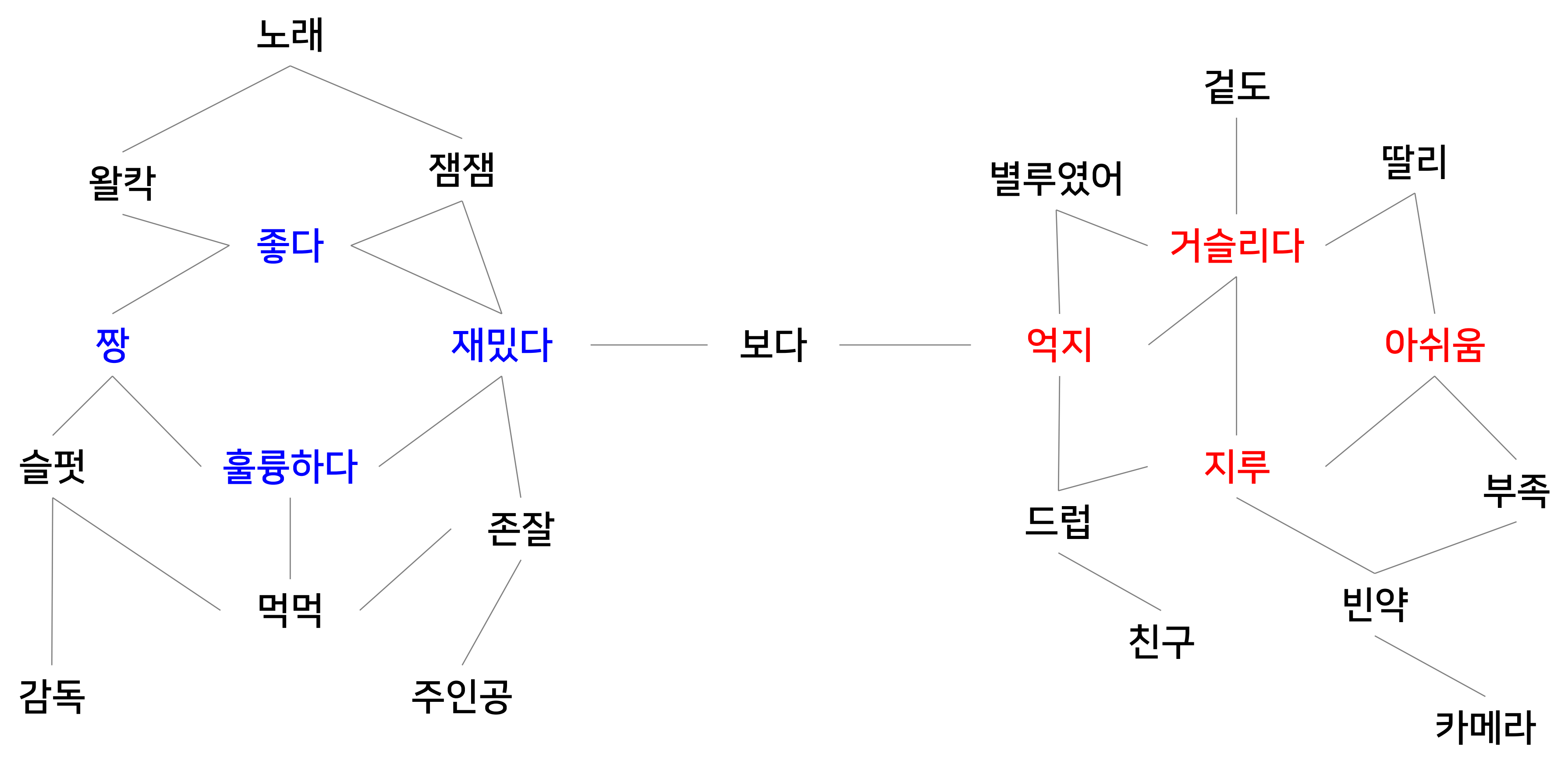
Graph-based SSL: Word Sentiment Propagation

- Labeled words

Positive pre-labeled words			Negative pre-labeled words		
좋다	역시	연기력	억지	지루	때우다
훌륭하다	즐겁다	아름답다	아쉬움	약하다	부담
재밌다	기대하다	긴장감	절대	흠	부끄럽다
사랑	좋아하다	매력	짜증	진부	고통
재미있다	멋지다	짱	지루함	싫다	난해
감동	대박	빠지다	밉다	별루	거슬리다
기대	웃다	멋있다	이상하다	심하다	망치다
최고	눈물	울다	킬링타임	식상하다	애매
재미	추천	코믹	필요없다	까다	거지
괜찮다	슬프다	오랜만	어이	졸리다	질질

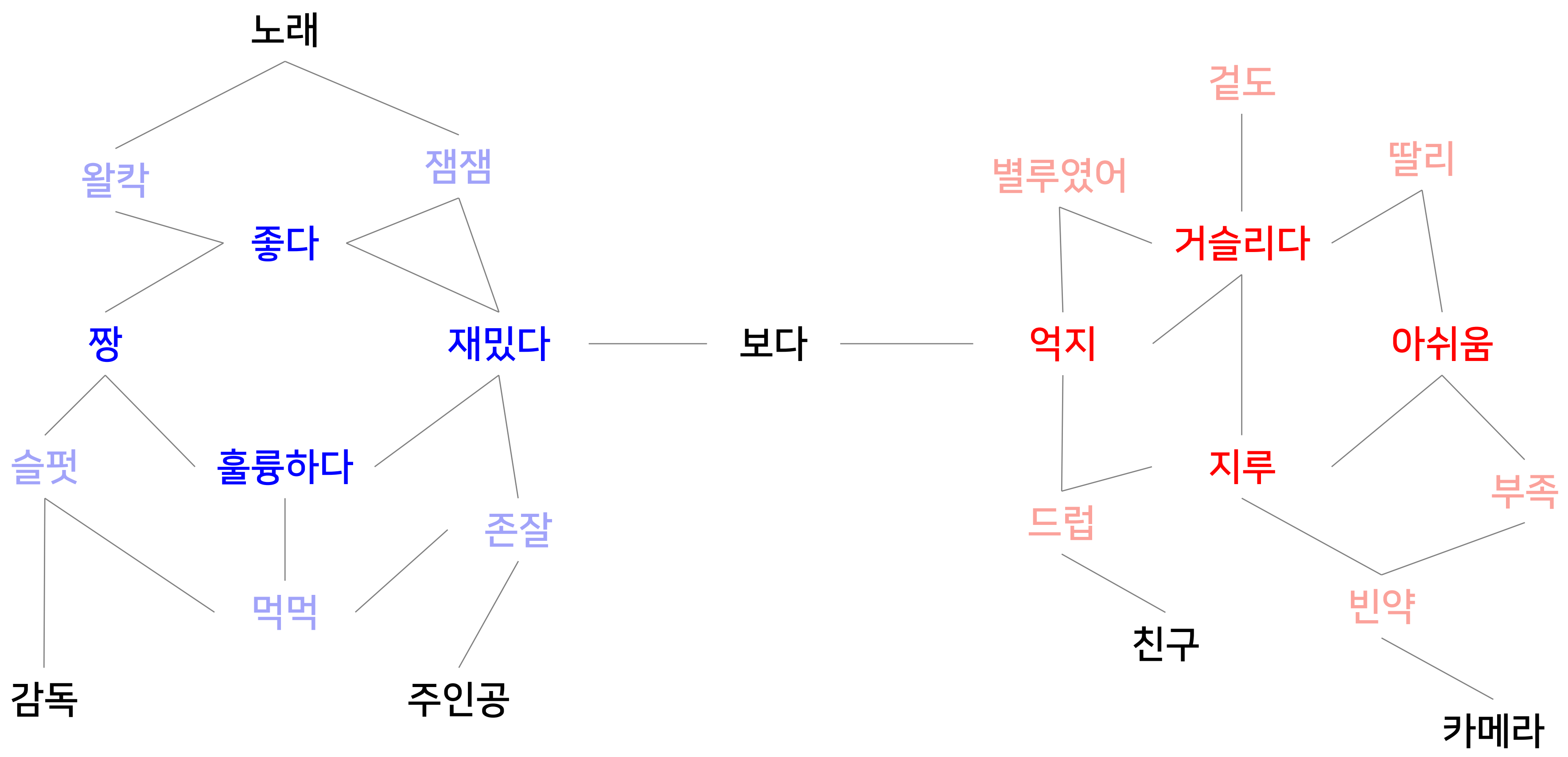
Graph-based SSL: Word Sentiment Propagation

- Label propagation based on Graph-based SSL



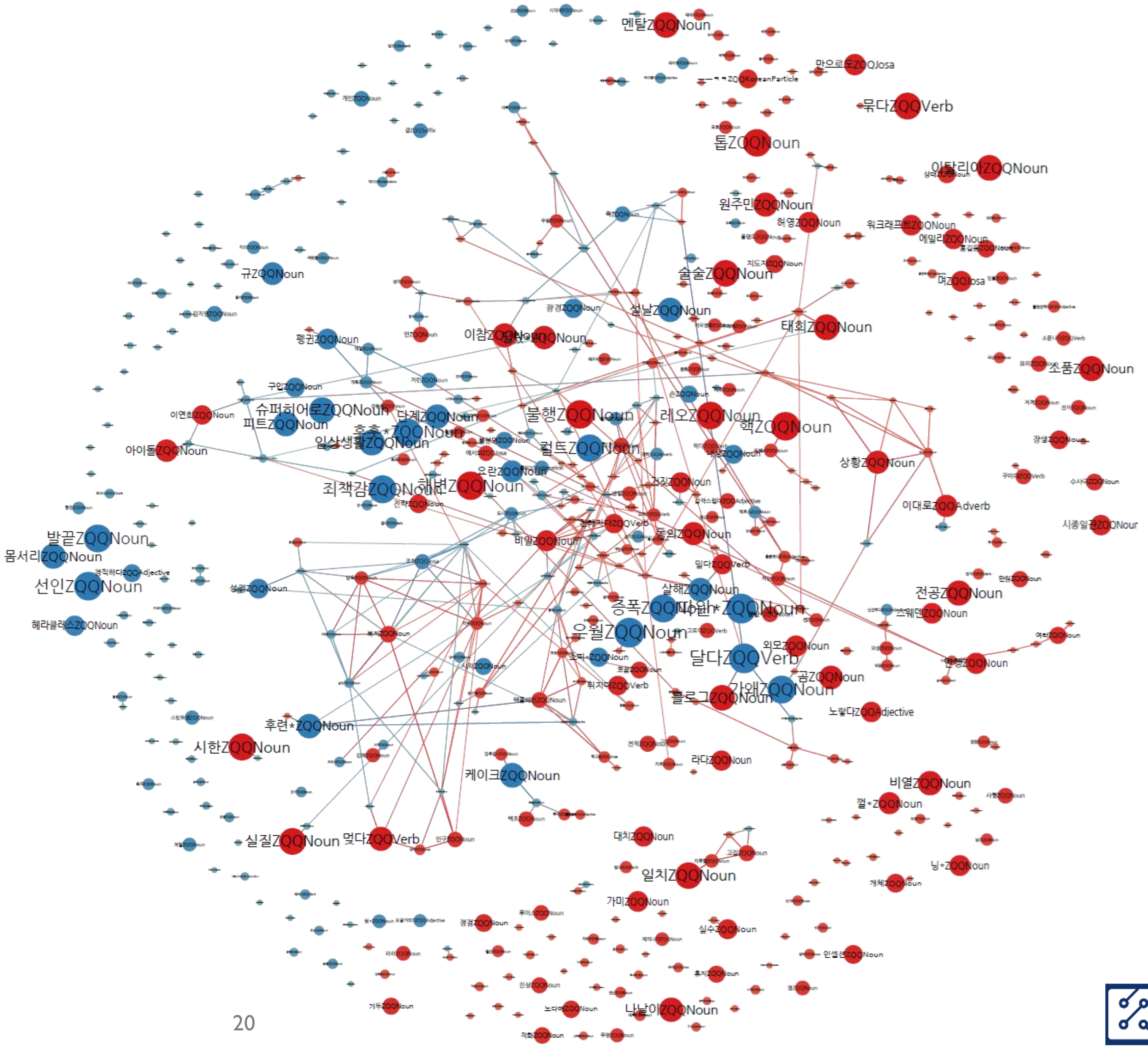
Graph-based SSL: Word Sentiment Propagation

- Label propagation based on Graph-based SSL



Graph-based SSL: Word Sentiment Propagation

- Propagated sentiment



Graph-based SSL: Word Sentiment Propagation

- Propagated sentiment

Positive		Negative	
Words	**Sentiment Score**	**Words**	**Sentiment Score**
슬펏	1	딸리	-0.9576
먹먹	1	별루였어	-0.5458
이뿌	0.8282	겉도	-0.4673
괜찮	0.5956	드럽	-0.4140
잼잼	0.5126	부족	-0.3713
신나요	0.4357	심해	-0.3449
눈시울	0.4089	어설퍼	-0.3123
짱	0.4002	미흡	-0.3062
왈칵	0.4001	지루	-0.2728
존잘	0.3851	빈약	-0.2531


Q
&
A

References

Research Papers

- Ben-David, S., Lu, T., & Pál, D. (2008, July). Does Unlabeled Data Provably Help? Worst-case Analysis of the Sample Complexity of Semi-Supervised Learning. In *COLT* (pp. 33-44).
- Bennett, K., & Demiriz, A. (1999). Semi-supervised support vector machines. *Advances in Neural Information processing systems*, 368-374.
- Blum, A., & Mitchell, T. (1998, July). Combining labeled and unlabeled data with co-training. In *Proceedings of the eleventh annual conference on Computational learning theory* (pp. 92-100). ACM.
- Fox-Roberts, P., & Rosten, E. (2014). Unbiased generative semi-supervised learning. *The Journal of Machine Learning Research*, *15*(1), 367-443.
- Jia, J., Schaub, M. T., Segarra, S., & Benson, A. R. (2019, July). Graph-based semi-supervised & active learning for edge flows. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 761-771).
- Kim, D., Seo, D., Cho, S., & Kang, P. (2017+). Multi-co-training for document classification using various document representations: TF-IDF, LDA, and Doc2Vec. under review.
- Kingma, D. P., Mohamed, S., Rezende, D. J., & Welling, M. (2014). Semi-supervised learning with deep generative models. In *Advances in Neural Information Processing Systems* (pp. 3581-3589).
- Singh, A., Nowak, R., & Zhu, X. (2009). Unlabeled data: Now it helps, now it doesn't. In *Advances in neural information processing systems* (pp. 1513-1520).
- Yu, S., Krishnapuram, B., Rosales, R., & Rao, R. B. (2011). Bayesian co-training. *The Journal of Machine Learning Research*, *12*, 2649-2680
- Zhou, Z. H., & Li, M. (2005, July). Semi-Supervised Regression with Co-Training. In IJCAI (Vol. 5, pp. 908-913).

References

<u>Other materials</u>

- Figures in the first page: 하상욱 단편시집 – 서울 시
- Zhu, X. (2007). Semi-Supervised Learning Tutorial. International Conference on Machine Learning (ICML 2007).
- Choi, S. (2015). Deep Learning: A Quick Overview. Deep Learning Workship. KIISE.
- Zien, A. (2008). Semi-Supervised Learning. Summer School on Neural Networks.
- Zhu, X. (2009). Tutorial on Semi-Supervised Learning. Theory and Practice of Computational Learning.