Computational and Methodological Challenges of Causal Inference in Networks

Panos Toulis

Harvard University, Statistics

May 2016

Introduction

- Randomization-based tests form an important class of statistical procedures.
- e.g., Fisher's sharp null for every unit i

$$Y_i(1) = Y_i(0),$$

where we have assumed *no interference*, i.e.,

$$Y_i(Z) = Y_i(Z') \text{ if } Z_i = Z'_i,$$

and thus $Y_i(Z) \equiv Y_i(Z_i)$.

- In the classical test, *imputation* of missing outcomes $Y_i(Z)$ (i.e., for Z not realized in the experiment) is straightforward.
- This imputation is **necessary** in a randomization-based method where we repeatedly sample Z and compute a test statistic.

Introduction

- Randomization-based tests form an important class of statistical procedures.
- e.g., Fisher's sharp null for every unit i

$$Y_i(1) = Y_i(0),$$

where we have assumed *no interference*, i.e.,

$$Y_i(Z) = Y_i(Z') \text{ if } Z_i = Z'_i,$$

and thus $Y_i(Z) \equiv Y_i(Z_i)$.

- In the classical test, *imputation* of missing outcomes $Y_i(Z)$ (i.e., for Z not realized in the experiment) is straightforward.
- This imputation is **necessary** in a randomization-based method where we repeatedly sample Z and compute a test statistic.

Introduction

- When there is interference things get complicated.
- Suppose G is undirected network of pretreatment connections and that unit i can only be affected by units $\{j : g_{ij} = 1\}$, i.e., assume *no foreign interference.*
- Some complications:
  (a) What is the treatment?

  $$Y_i(Z_i, \#\text{neighbors treated}), \text{ or } Y_i(Z_i, \%\text{neighbors treated})?$$

  (b) What is the null hypothesis H_0? for example,

  $$Y_i(Z_i, > 0 \text{ neighbors treated}) = Y_i(Z_i, 0 \text{ neighbors treated}).$$

  (c) How to randomize and how to analyze? What is the role of the network G?

Introduction

- When there is interference things get complicated.
- Suppose G is undirected network of pretreatment connections and that unit i can only be affected by units $\{j : g_{ij} = 1\}$, i.e., assume *no foreign interference.*
- Some complications:
  (a) What is the treatment?

$$Y_i(Z_i, \#\text{neighbors treated}), \text{ or } Y_i(Z_i, \%\text{neighbors treated})?$$

  (b) What is the null hypothesis H_0? for example,

$$Y_i(Z_i, > 0 \text{ neighbors treated}) = Y_i(Z_i, 0 \text{ neighbors treated}).$$

  (c) How to randomize and how to analyze? What is the role of the network G?

Introduction

- When there is interference things get complicated.
- Suppose G is undirected network of pretreatment connections and that unit i can only be affected by units $\{j : g_{ij} = 1\}$, i.e., assume *no foreign interference.*
- Some complications:
  (a) What is the treatment?

$$Y_i(Z_i, \#\text{neighbors treated}), \text{ or } Y_i(Z_i, \%\text{neighbors treated})?$$

  (b) What is the null hypothesis H_0? for example,

$$Y_i(Z_i, > 0 \text{ neighbors treated}) = Y_i(Z_i, 0 \text{ neighbors treated}).$$

  (c) How to randomize and how to analyze? What is the role of the network G?

Introduction

- When there is interference things get complicated.
- Suppose G is undirected network of pretreatment connections and that unit i can only be affected by units $\{j : g_{ij} = 1\}$, i.e., assume *no foreign interference.*
- Some complications:
  (a) What is the treatment?

$$Y_i(Z_i, \#\text{neighbors treated}), \text{ or } Y_i(Z_i, \%\text{neighbors treated})?$$

  (b) What is the null hypothesis H_0? for example,

$$Y_i(Z_i, > 0 \text{ neighbors treated}) = Y_i(Z_i, 0 \text{ neighbors treated}).$$

  (c) How to randomize and how to analyze? What is the role of the network G?

Bowers et.al. [3]

- Defines a causal model H as a map from $\mathbf{Y}(0)$ to any $\mathbf{Y}(Z)$:

$$\mathbf{Y}(Z) = H(\mathbf{Y}(0), Z; \theta)$$

  - Can generate **all** missing $\mathbf{Y}(Z)$ in a randomization-based test.
- Parameter $\theta = (\beta, \tau)$ captures main treatment effect (β) and interference (τ) effect.
- Treatment of i depends on immediate neighborhood (generally prevalent assumption).
- Null hypothesis is, for example, $\tau = 0$.
- Computation is an issue when deciding the test statistic (KS divergence between treated-control unit outcomes in this paper).

Bowers et.al. [3]

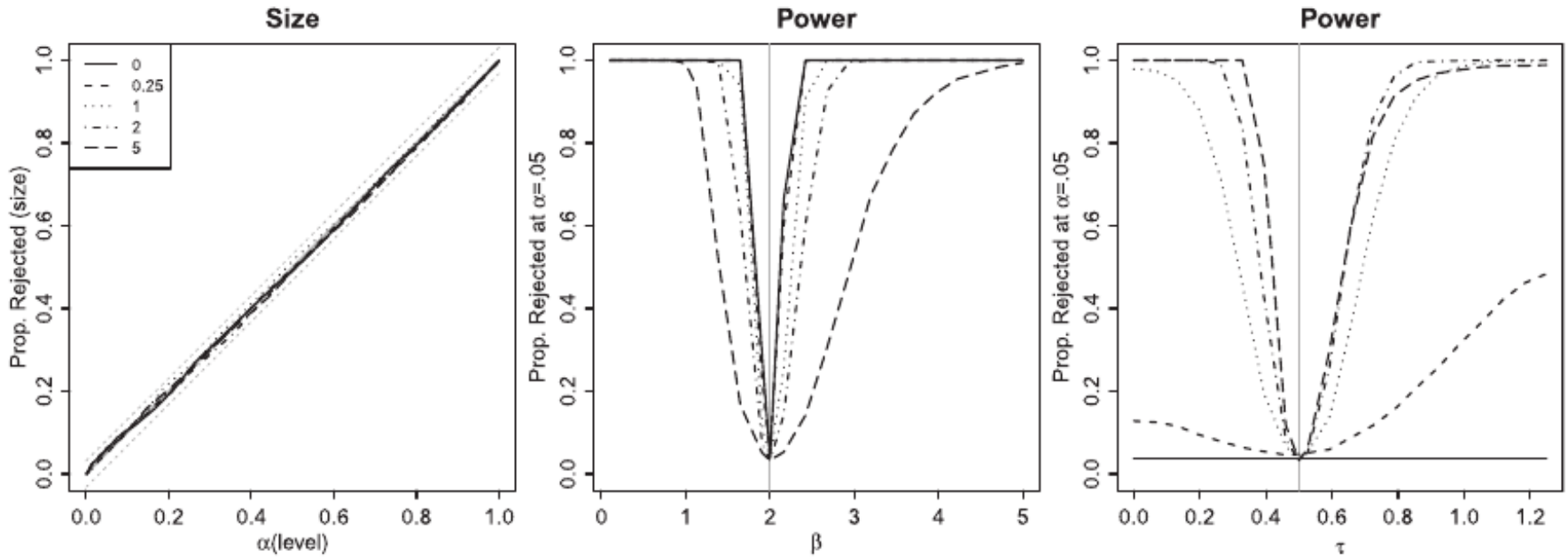
- Defines a causal model H as a map from $\mathbf{Y}(0)$ to any $\mathbf{Y}(Z)$:

$$\mathbf{Y}(Z) = H(\mathbf{Y}(0), Z; \theta)$$

  - Can generate **all** missing $\mathbf{Y}(Z)$ in a randomization-based test.
- Parameter $\theta = (\beta, \tau)$ captures main treatment effect (β) and interference (τ) effect.
- Treatment of i depends on immediate neighborhood (generally prevalent assumption).
- Null hypothesis is, for example, $\tau = 0$.
- Computation is an issue when deciding the test statistic (KS divergence between treated-control unit outcomes in this paper).

Bowers et.al. [3]

- Interestingly, the network topology affects how the test performs.
  - dense graphs give more information for interference; sparse graphs give information for main effect (also in [4]).



Ugander et.al.[6]

- Randomization is hard because of network dependencies.
- One idea is to cluster the network and then assign treatments on the cluster level.
- Use Horvitz-Thompson estimator to estimate average causal effects (as prescribed by Aronow [1]).
- Calculation of exposure probabilities is crucial.

Ugander et.al.[6]

- Randomization is hard because of network dependencies.
- One idea is to cluster the network and then assign treatments on the cluster level.
- Use Horvitz-Thompson estimator to estimate average causal effects (as prescribed by Aronow [1]).
- Calculation of exposure probabilities is crucial.

Ugander et.al.[6]

1. Probability of exposure in clustered randomization can be obtained sometimes through dynamic programming.
2. **Example:** Fix unit i and let $f(j,k)$ be the probability that i will get at least k treated neighbors if we treat clusters $\{1,2,..,j\}$ and let $c_{ij} = \#$connections to unit i from cluster j; then

$$f(j,k) = \underbrace{p}_{Pr(j \text{ treated})} f(j-1, k-c_{ij}) + (1-p) f(j-1,k).$$

3. Computation complexity is

$$O(\text{Max\_Degree} \cdot \#\text{Clusters}).$$

4. (Theorem) If $\mathbf{Y}(Z)$, Max_Degree, and #Clusters are $O(1)$ then the variance of the HT estimator is $O(1/\#\text{Units})$.

Ugander et.al.[6]

1. Probability of exposure in clustered randomization can be obtained sometimes through dynamic programming.
2. **Example:** Fix unit i and let $f(j,k)$ be the probability that i will get at least k treated neighbors if we treat clusters $\{1,2,..,j\}$ and let $c_{ij} = \#$connections to unit i from cluster j; then

$$f(j,k) = \underbrace{p}_{Pr(j \text{ treated})} f(j-1, k-c_{ij}) + (1-p) f(j-1,k).$$

3. Computation complexity is

$$O(\text{Max\_Degree} \cdot \#\text{Clusters}).$$

4. (Theorem) If $\mathbf{Y}(Z)$, Max_Degree, and #Clusters are $O(1)$ then the variance of the HT estimator is $O(1/\#\text{Units})$.

Ugander et.al.[6]

1. Probability of exposure in clustered randomization can be obtained sometimes through dynamic programming.
2. **Example:** Fix unit i and let $f(j,k)$ be the probability that i will get at least k treated neighbors if we treat clusters $\{1,2,..,j\}$ and let $c_{ij} = \#$connections to unit i from cluster j; then

$$f(j,k) = \underbrace{p}_{Pr(j \text{ treated})} f(j-1, k-c_{ij}) + (1-p) f(j-1,k).$$

3. Computation complexity is

$$O(\text{Max\_Degree} \cdot \#\text{Clusters}).$$

4. (Theorem) If $\mathbf{Y}(Z)$, Max_Degree, and #Clusters are $O(1)$ then the variance of the HT estimator is $O(1/\#\text{Units})$.

Ugander et.al.[6]

1. Probability of exposure in clustered randomization can be obtained sometimes through dynamic programming.
2. **Example:** Fix unit i and let $f(j,k)$ be the probability that i will get at least k treated neighbors if we treat clusters $\{1,2,..,j\}$ and let $c_{ij} = \#$connections to unit i from cluster j; then

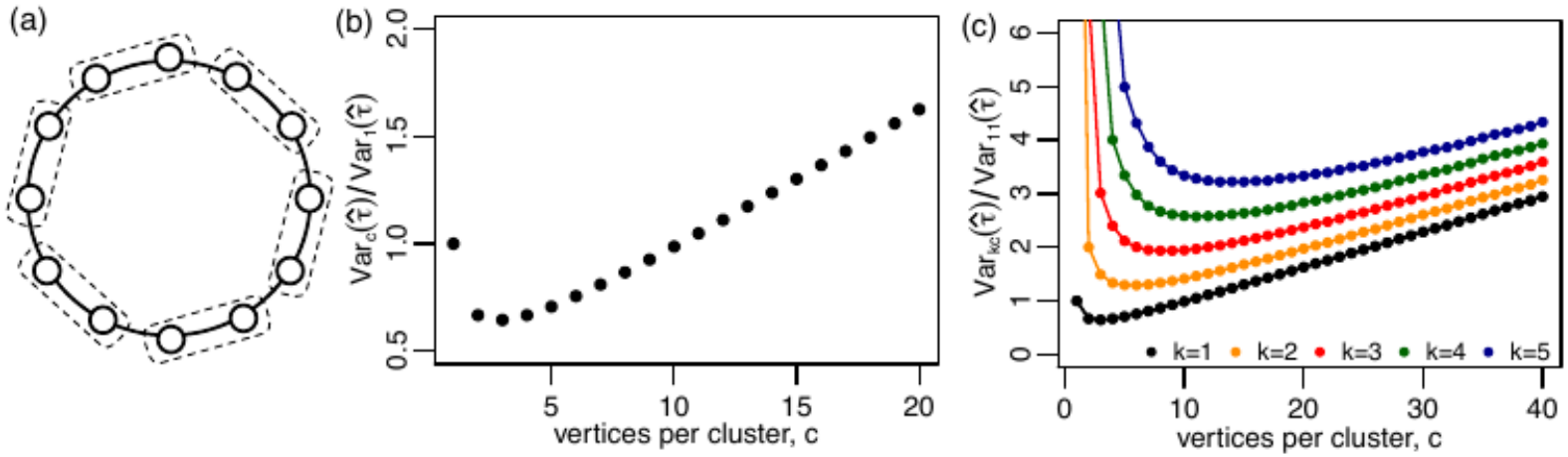
$$f(j,k) = \underbrace{p}_{Pr(j \text{ treated})} f(j-1, k-c_{ij}) + (1-p) f(j-1,k).$$

3. Computation complexity is

$$O(\text{Max\_Degree} \cdot \#\text{Clusters}).$$

4. (Theorem) If $\mathbf{Y}(Z)$, Max_Degree, and #Clusters are $O(1)$ then the variance of the HT estimator is $O(1/\#\text{Units})$.

Ugander et.al.[6]: Example



- Assume no foreign interference such that

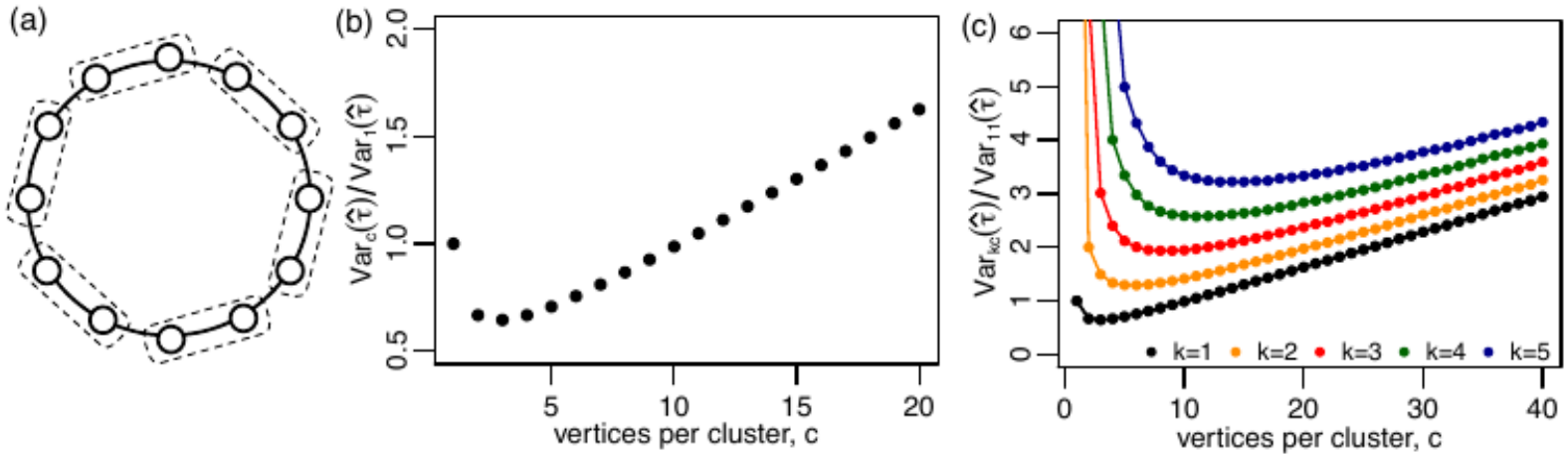
$$Y_i(Z) \equiv Y_i(Z_i, Z_{\text{neighbor 1}}, Z_{\text{neighbor 2}})$$

- Estimate

$$Ave(Y_i(1,1,1)) - Ave(Y_i(0,0,0))$$

- Suppose c is the cluster size. Variance of HT estimator has $\propto 1/c$ term (between-cluster var.) and $\propto c$ term (within-cluster var.) Optimal when $c = 3$.

Ugander et.al.[6]: Example



- Assume no foreign interference such that

$$Y_i(Z) \equiv Y_i(Z_i, Z_{\text{neighbor 1}}, Z_{\text{neighbor 2}})$$

- Estimate

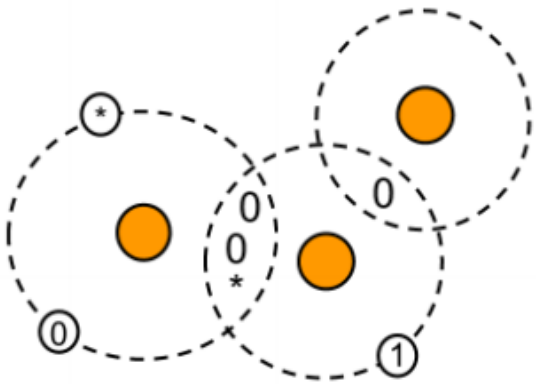
$$Ave(Y_i(1,1,1)) - Ave(Y_i(0,0,0))$$

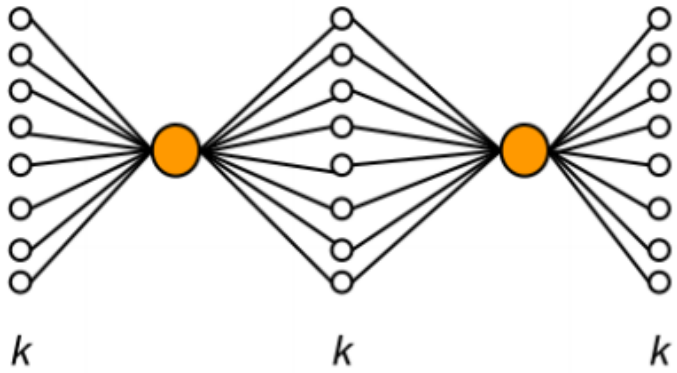
- Suppose c is the cluster size. Variance of HT estimator has $\propto 1/c$ term (between-cluster var.) and $\propto c$ term (within-cluster var.) **Optimal when** $c = 3$.

T. and Kao [5]

- Suppose we wish to estimate the effect of k neighbors (right figure). For this, only two possible assignments are informative.
- In *insulated neighbors randomization* one sets the shared neighborhood in control before assigning composite treatments (e.g., entire neighborhood treatments) as in left figure.
- Simple bias characterization based on shared neighbors (SN):

$$\text{bias} = \%\ \text{SN} \times [\text{effect from SN - non-SN}].$$

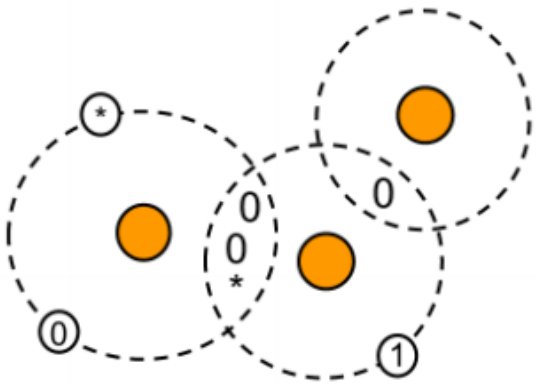


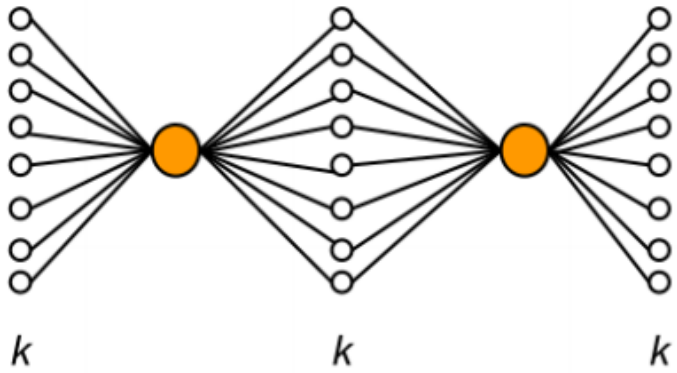


T. and Kao [5]

- Suppose we wish to estimate the effect of k neighbors (right figure). For this, only two possible assignments are informative.
- In *insulated neighbors randomization* one sets the shared neighborhood in control before assigning composite treatments (e.g., entire neighborhood treatments) as in left figure.
- Simple bias characterization based on shared neighbors (SN):

$$\text{bias} = \%\ \text{SN} \times [\text{effect from SN - non-SN}].$$





Athey et. al. [2]

Also focused on testing a hypothesis H_0 such as "no spillover effect".
General approach:

- (Theorem) H_0 defines a partition of the space of assignments.
- Select set of *focal units*.
- Compute set of assignments for which we can impute the missing outcomes of the focal units under H_0.
- Compute test statistic using outcomes of the focal units.
- Execute randomization test **conditional** on the partition that the realized assignment is a member of.

Athey et. al. [2]

Devil is in the details: How to choose the focal units?

- Goal is to maximize the power of the test.
- The power depends on
  (a) alternative hypothesis
  (b) choice of statistic
  (c) network structure
- (Apparently) hard graph-theoretic problem; e.g., maximize focal-nonfocal edges:

$$\text{focal units} = \arg\max_{S} \sum_{i \in S, j \notin S} g_{ij}.$$

Athey et. al. [2]

Devil is in the details: How to choose the focal units?

- Goal is to maximize the power of the test.
- The power depends on
  (a) alternative hypothesis
  (b) choice of statistic
  (c) network structure
- (Apparently) hard graph-theoretic problem; e.g., maximize focal-nonfocal edges:

$$\text{focal units} = \arg\max_{S} \sum_{i \in S, j \notin S} g_{ij}.$$

Athey et. al. [2]

Devil is in the details: How to choose the focal units?

- Goal is to maximize the power of the test.
- The power depends on
  (a) alternative hypothesis
  (b) choice of statistic
  (c) network structure
- (Apparently) hard graph-theoretic problem; e.g., maximize focal-nonfocal edges:

$$\text{focal units} = \arg\max_{S} \sum_{i \in S, j \notin S} g_{ij}.$$

References

[1] Peter M Aronow and Cyrus Samii. Estimating average causal effects under general interference. In *Summer Meeting of the Society for Political Methodology, University of North Carolina, Chapel Hill, July*, pages 19–21. Citeseer, 2012.

[2] Susan Athey, Dean Eckles, and Guido Imbens. Exact p-values for network interference. *arxiv*, 20125.

[3] Jake Bowers, Mark M Fredrickson, and Costas Panagopoulos. Reasoning about interference between units: A general framework. *Political Analysis*, page mps038, 2012.

[4] Panos Toulis and Edward Kao. Estimation of causal peer influence effects. In *Proceedings of The 30th International Conference on Machine Learning*, pages 1489–1497, 2013.

[5] Panos Toulis and Edward Kao. Estimation of causal peer influence effects. In *Proceedings of The 30th International Conference on Machine Learning*, pages 1489–1497, 2013.

[6] Johan Ugander, Brian Karrer, Lars Backstrom, and Jon Kleinberg. Graph cluster randomization: Network exposure to multiple universes. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 329–337. ACM, 2013.