Local cliques in ER-perturbed random geometric graphs

Minghao Tian

The Ohio State University

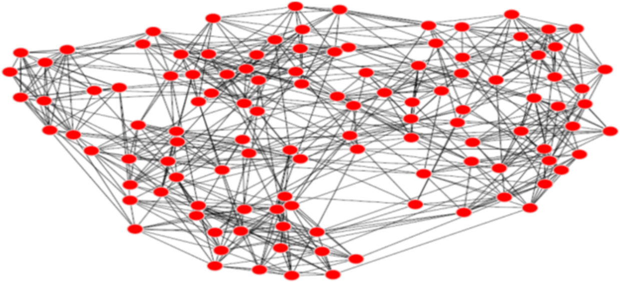
Joint work with Matthew Kahle and Yusu Wang

full version: https://arxiv.org/abs/1810.08383

November 6, 2018

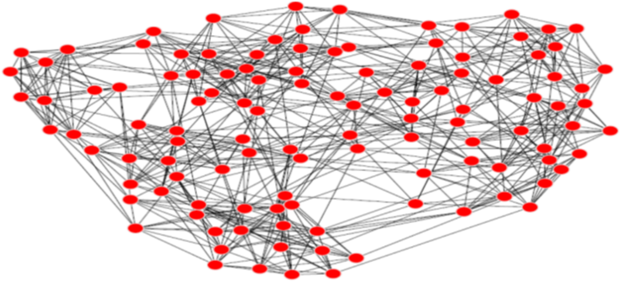
Introduction

- Graph / Network — common data type



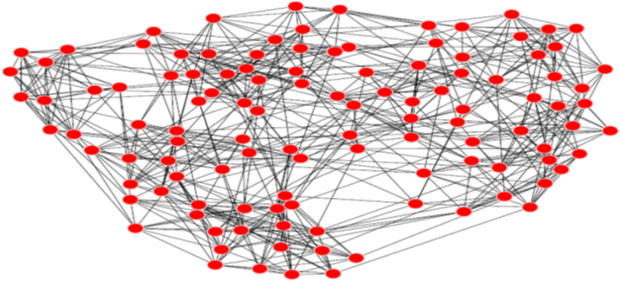
Introduction

- Graph / Network — common data type
- Often an input graph G can be viewed as a noisy observation (perturbed version) of a hidden ground truth graph G^*



Introduction

- Graph / Network — common data type
- Often an input graph G can be viewed as a noisy observation (perturbed version) of a hidden ground truth graph G^*
- High level goal:
  - Inference about true graph G^*, or analyze properties of perturbed graphs



Our Network Model

The true graph $G^* = (V, E^*)$ is a random geometric graph(RGG) where:

- $V = V_n$ are n points sampled *i.i.d* from a probability density function induced by a "nice" measure μ on a "nice" metric space $\mathcal{M} = (M, d)$

Our Network Model

The true graph $G^* = (V, E^*)$ is a random geometric graph(RGG) where:

- $V = V_n$ are n points sampled *i.i.d* from a probability density function induced by a "nice" measure μ on a "nice" metric space $\mathcal{M} = (M, d)$
- $E^* = E^*_{r,n} = \{(u, v) \mid d_M(u, v) \leq r, u, v \in V\}$ is the $r-$neighborhood graph for some parameter $r > 0$

Our Network Model

The true graph $G^* = (V, E^*)$ is a random geometric graph(RGG) where:

- $V = V_n$ are n points sampled *i.i.d* from a probability density function induced by a "nice" measure μ on a "nice" metric space $\mathcal{M} = (M, d)$
- $E^* = E^*_{r,n} = \{(u, v) \mid d_M(u, v) \le r, u, v \in V\}$ is the $r-$neighborhood graph for some parameter $r > 0$

The observed graph G is an Erdős-Rényi type(ER) (p, q)-perturbation of G^* where:

Our Network Model

The true graph $G^* = (V, E^*)$ is a random geometric graph(RGG) where:

- $V = V_n$ are n points sampled *i.i.d* from a probability density function induced by a "nice" measure μ on a "nice" metric space $\mathcal{M} = (M, d)$
- $E^* = E^*_{r,n} = \{(u, v) \mid d_M(u, v) \le r, u, v \in V\}$ is the $r-$neighborhood graph for some parameter $r > 0$

The observed graph G is an Erdős-Rényi type(ER) (p, q)-perturbation of G^* where:

- $p-$**deletion:** For each existing edge $(u, v) \in E^*$, we delete edge (u, v) with probability p

Our Network Model

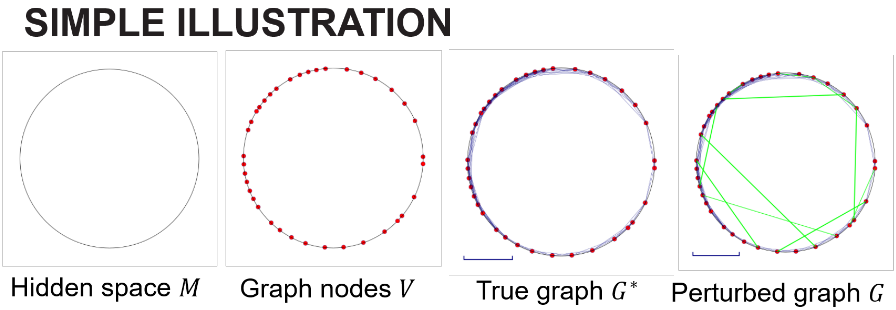
The true graph $G^* = (V, E^*)$ is a random geometric graph(RGG) where:

- $V = V_n$ are n points sampled *i.i.d* from a probability density function induced by a "nice" measure μ on a "nice" metric space $\mathcal{M} = (M, d)$
- $E^* = E^*_{r,n} = \{(u, v) \mid d_M(u, v) \leq r, u, v \in V\}$ is the $r-$neighborhood graph for some parameter $r > 0$

The observed graph G is an Erdős-Rényi type(ER) (p, q)-perturbation of G^* where:

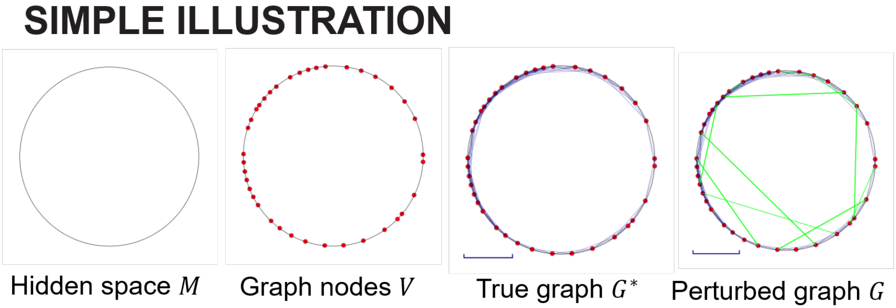
- $p-$**deletion:** For each existing edge $(u, v) \in E^*$, we delete edge (u, v) with probability p
- $q-$**insertion:** For each non-existent edge $(u, v) \notin E^*$, we insert edge (u, v) with probability q

Our Network Model



[1] Ronald Meester and Rahul Roy. Continuum percolation, volume 119. Cambridge University Press, 1996.

Our Network Model



Remark: Our model is related to the continuum percolation theory[1].

[1] Ronald Meester and Rahul Roy. Continuum percolation, volume 119. Cambridge University Press, 1996.

Clique Number

The clique number of a graph $\mathcal{G}$, denoted $\omega(\mathcal{G})$, is the number of vertices in a maximum clique of $\mathcal{G}$.

Clique Number

The clique number of a graph $\mathcal{G}$, denoted $\omega(\mathcal{G})$, is the number of vertices in a maximum clique of $\mathcal{G}$.

- In Erdős-Rényi graphs $G(n, p)$:
  - If $0 < p < 1$ is a constant, then $\lim_{n\to\infty} Pr[\omega(G(n,p)) = k(n) \text{ or } k(n)-1] = 1$, where $k(n) \sim 2\log_{1/p} n$ (The celebrated two-point concentration[2])

[1] For the explicit expression of $k(n)$, check: Béla Bollobás and Paul Erdős. Cliques in random graphs. In Mathematical Proceedings of the Cambridge Philosophical Society, volume 80, pages 419 427. Cambridge University Press, 1976.

Noga Alon and Joel H. Spencer. The Probabilistic Method. Wiley Publishing, 4th edition, 2016.

Clique Number

The clique number of a graph $\mathcal{G}$, denoted $\omega(\mathcal{G})$, is the number of vertices in a maximum clique of $\mathcal{G}$.

- In Erdős-Rényi graphs $G(n,p)$:
  - If $0 < p < 1$ is a constant, then $\lim_{n\to\infty} Pr[\omega(G(n,p)) = k(n) \text{ or } k(n)-1] = 1$, where $k(n) \sim 2\log_{1/p} n$ (The celebrated two-point concentration[2])
- In standard random geometric graphs(the underlying space is $\mathbb{R}^d$):
  - Has dramatically different behaviors when different ranges of r are chosen [3].

[1]For the explicit expression of $k(n)$, check: Béla Bollobás and Paul Erdős. Cliques in random graphs. In Mathematical Proceedings of the Cambridge Philosophical Society, volume 80, pages 419 427. Cambridge University Press, 1976.

Noga Alon and Joel H. Spencer. The Probabilistic Method. Wiley Publishing, 4th edition, 2016.

[3]Colin McDiarmid and Tobias Müller. On the chromatic number of random geometric graphs. Combinatorica, 31(4):423488, 2011.

Clique Number

The clique number of a graph $\mathcal{G}$, denoted $\omega(\mathcal{G})$, is the number of vertices in a maximum clique of $\mathcal{G}$.

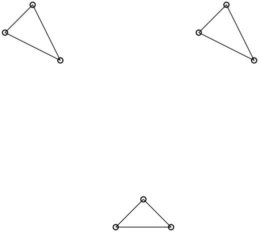
- In Erdős-Rényi graphs $G(n, p)$:
  - If $0 < p < 1$ is a constant, then $\lim_{n \to \infty} Pr[\omega(G(n,p)) = k(n) \text{ or } k(n) - 1] = 1$, where $k(n) \sim 2 \log_{1/p} n$ (The celebrated two-point concentration[2])
- In standard random geometric graphs(the underlying space is $\mathbb{R}^d$):
  - Has dramatically different behaviors when different ranges of r are chosen [3].
- In our model?

[1] For the explicit expression of $k(n)$, check: Béla Bollobás and Paul Erdős. Cliques in random graphs. In Mathematical Proceedings of the Cambridge Philosophical Society, volume 80, pages 419 427. Cambridge University Press, 1976.

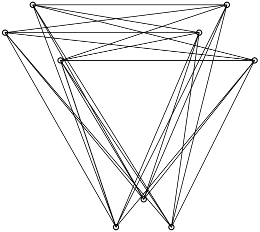
Noga Alon and Joel H. Spencer. The Probabilistic Method. Wiley Publishing, 4th edition, 2016.

[3] Colin McDiarmid and Tobias Müller. On the chromatic number of random geometric graphs. Combinatorica, 31(4):423488, 2011.

The Clique Number of The Union Graph



$\mathcal{G} = 3\ K_3$'s $(\omega(\mathcal{G}) = 3)$



The complement $\mathcal{G}^c$ $(\omega(\mathcal{G}^c) = 3)$

The union graph is a K_9 whose clique number is 9.
(consider $\sqrt{n}\ K_{\sqrt{n}}$ and its complement)

Edge Clique Number

The edge clique number of an edge e in $\mathcal{G}$, denoted $\omega_{\mathcal{G}}(e)$, is the number of vertices in a maximum clique in $\mathcal{G}$ containing e.

Edge Clique Number

The edge clique number of an edge e in $\mathcal{G}$, denoted $\omega_{\mathcal{G}}(e)$, is the number of vertices in a maximum clique in $\mathcal{G}$ containing e.

- It is the "local version" of clique number.

Edge Clique Number

The edge clique number of an edge e in $\mathcal{G}$, denoted $\omega_{\mathcal{G}}(e)$, is the number of vertices in a maximum clique in $\mathcal{G}$ containing e.

- It is the "local version" of clique number.
- Intuitively, an edge in the observed graph G(ER-perturbed) can come from either the random geometric graph G^* or the Erdős-Rényi perturbation (inserted edges).

Edge Clique Number

The edge clique number of an edge e in $\mathcal{G}$, denoted $\omega_{\mathcal{G}}(e)$, is the number of vertices in a maximum clique in $\mathcal{G}$ containing e.

- It is the "local version" of clique number.
- Intuitively, an edge in the observed graph G(ER-perturbed) can come from either the random geometric graph G^* or the Erdős-Rényi perturbation (inserted edges).
- (Main result) The edge clique number exhibits fundamentally different behaviors for these two types of edges.

"Nice" Measure μ and "Nice" Metric Space $\mathcal{M}$

“Nice” Measure μ and “Nice” Metric Space $\mathcal{M}$

- $\mathcal{M}$ — a compact geodesic metric space

"Nice" Measure μ and "Nice" Metric Space $\mathcal{M}$

- $\mathcal{M}$ — a compact geodesic metric space
  - Any two points in it are connected by a path whose length equals the distance between them. Uniqueness of geodesics is not required.

“Nice” Measure μ and “Nice” Metric Space $\mathcal{M}$

- $\mathcal{M}$ — a compact geodesic metric space
  - Any two points in it are connected by a path whose length equals the distance between them. Uniqueness of geodesics is not required.
  - e.g. Riemannian manifolds, path-connected compact sets in the Euclidean space, etc.

“Nice” Measure μ and “Nice” Metric Space $\mathcal{M}$

- $\mathcal{M}$ — a compact geodesic metric space
  - Any two points in it are connected by a path whose length equals the distance between them. Uniqueness of geodesics is not required.
  - e.g. Riemannian manifolds, path-connected compact sets in the Euclidean space, etc.
- μ — a doubling measure

“Nice” Measure μ and “Nice” Metric Space $\mathcal{M}$

- $\mathcal{M}$ — a compact geodesic metric space
  - Any two points in it are connected by a path whose length equals the distance between them. Uniqueness of geodesics is not required.
  - e.g. Riemannian manifolds, path-connected compact sets in the Euclidean space, etc.

- μ — a doubling measure
  - Every metric ball (with positive radius) has finite and positive measure and there is a constant $L = L(\mu)$ s.t. for all $x \in M$ and every $R > 0$, we have $\mu(B_{2R}(x)) \leq L \cdot \mu(B_R(x))$.

"Nice" Measure μ and "Nice" Metric Space $\mathcal{M}$

- $\mathcal{M}$ — a compact geodesic metric space
  - Any two points in it are connected by a path whose length equals the distance between them. Uniqueness of geodesics is not required.
  - e.g. Riemannian manifolds, path-connected compact sets in the Euclidean space, etc.
- μ — a doubling measure
  - Every metric ball (with positive radius) has finite and positive measure and there is a constant $L = L(\mu)$ s.t. for all $x \in M$ and every $R > 0$, we have $\mu(B_{2R}(x)) \leq L \cdot \mu(B_R(x))$.
  - L is the doubling constant and μ is an $L-$doubling measure.

Further Assumption on "Nice" Measure μ

For technical reasons, we need an assumption on the parameter r (for the RGG G^*), as well as a condition on the measure μ.

Assumption-A

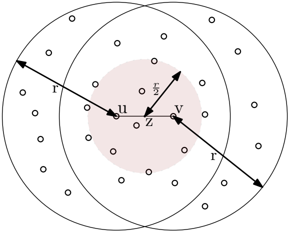
The parameter r and the doubling measure μ satisfy the following condition:

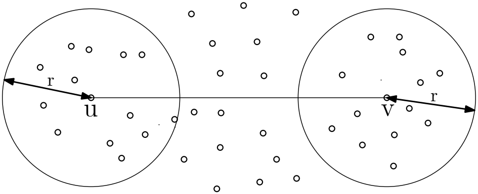
There exist $\mathrm{s} \geq \frac{13 \ln n}{n} \left(= \Omega(\frac{\ln n}{n})\right)$ and a constant ρ such that for any $x \in X$

(Density-cond) $\mu\left(B_{r/2}(x)\right) \geq \mathrm{s}$.

(Regularity-cond) $\mu\left(B_{r/2}(x)\right) \leq \rho \mathrm{s}$

Two Types of Edges — Good Edges and Bad Edges





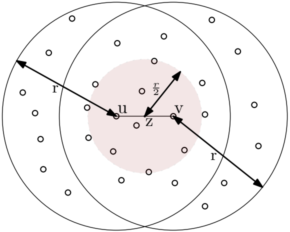
Good edge: $d(u, v) \leq r$

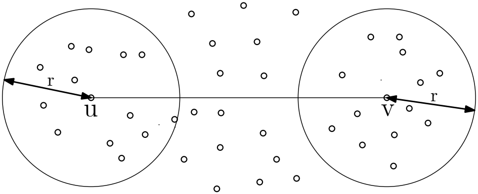
(Edges from RGG)

Bad edge: $\forall x \in N_{G^*}(u), y \in N_{G^*}(v)$, $d(x, y) > r$.

(Edges from ER perturbation)

Two Types of Edges — Good Edges and Bad Edges





Good edge: $d(u, v) \leq r$

(Edges from RGG)

Bad edge: $\forall x \in N_{G^*}(u), y \in N_{G^*}(v)$, $d(x, y) > r$.

(Edges from ER perturbation)

Remark: There are "not-so-bad" edges other than these two types.

Main Results

Theorem (Simplified, Insertion-only)

Let G^* be the true graph generated as described, and G a graph obtained after random $q-$insertion. Under Assumption-A, for any insertion probability $q = o(1)$, with high probability,

- for all good edges $e \in G$, we have $\omega_G(e) = \Theta(sn)$
- for all bad edges $e \in G$, we have $\omega_G(e) = o(sn)$

Main Results

Theorem (Simplified, Insertion-only)

Let G^* be the true graph generated as described, and G a graph obtained after random $q-$insertion. Under Assumption-A, for any insertion probability $q = o(1)$, with high probability,

- for all good edges $e \in G$, we have $\omega_G(e) = \Theta(sn)$
- for all bad edges $e \in G$, we have $\omega_G(e) = o(sn)$

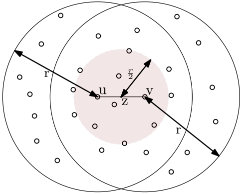
Theorem (Simplified, $(p, q)-$perturbation)

Let G^* be the true graph generated as described, and G a graph obtained after random $p-$deletion and $q-$insertion. Under Assumption-A and assume $sn = \Theta(\ln n)$, for any constant $p \in (0, 1)$ and $q = o\left(\left(\frac{1}{n}\right)^{\frac{c}{\ln \ln n}} \frac{\ln \ln n}{\ln n}\right)$, with high probability,

- for all good edges $e \in G$, we have $\omega_G(e) = \Omega(\ln \ln n)$
- for all bad edges $e \in G$, we have $\omega_G(e) = o(\ln \ln n)$

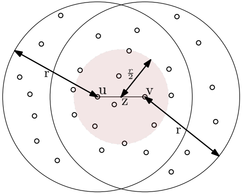
Sketch of Proof of the Insertion-only Case

- The lower bound for good edges



Sketch of Proof of the Insertion-only Case

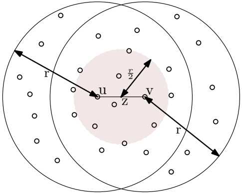
- The lower bound for good edges



  - $B_r(u) \cap B_r(v)$ must contain an $r/2$ centered at the midpoint z of a geodesic connecting u to v in M

Sketch of Proof of the Insertion-only Case

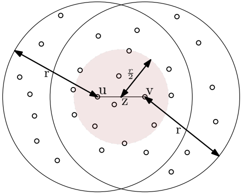
- The lower bound for good edges



  - $B_r(u) \cap B_r(v)$ must contain an $r/2$ centered at the midpoint z of a geodesic connecting u to v in M
  - At most $\binom{n}{2}$ such $r/2-$balls

Sketch of Proof of the Insertion-only Case

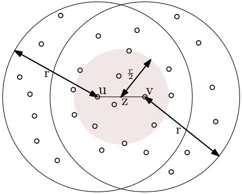
- The lower bound for good edges



  - $B_r(u) \cap B_r(v)$ must contain an $r/2$ centered at the midpoint z of a geodesic connecting u to v in M
  - At most $\binom{n}{2}$ such $r/2-$balls
  - The number of points in any $r/2-$ball can be bounded from below with **very** high probability by applying the Chernoff bound

Sketch of Proof of the Insertion-only Case

- The lower bound for good edges

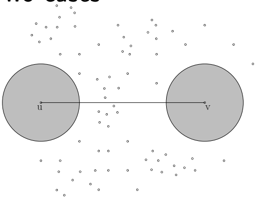


  - $B_r(u) \cap B_r(v)$ must contain an $r/2$ centered at the midpoint z of a geodesic connecting u to v in M
  - At most $\binom{n}{2}$ such $r/2-$balls
  - The number of points in any $r/2-$ball can be bounded from below with **very** high probability by applying the Chernoff bound
  - Union bound

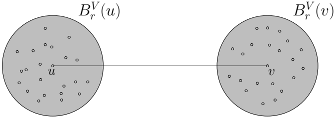
Sketch of Proof of the Insertion-only Case (cont'd)

- The upper bound for bad edges

- The upper bound for bad edges
  - Two cases



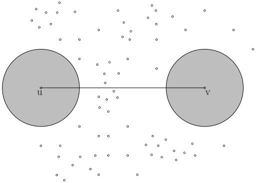
Case (a) — $\tilde{A}_{uv}$

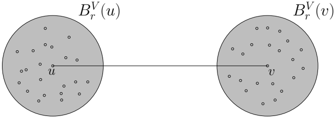


Case (b) — B_{uv}

Sketch of Proof of the Insertion-only Case (cont'd)

- The upper bound for bad edges
  - Two cases



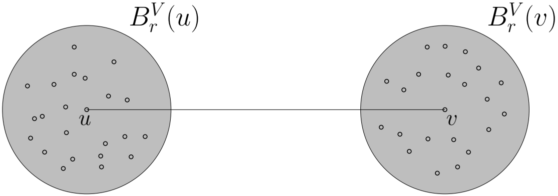


Case (a) — $\tilde{A}_{uv}$ Case (b) — B_{uv}

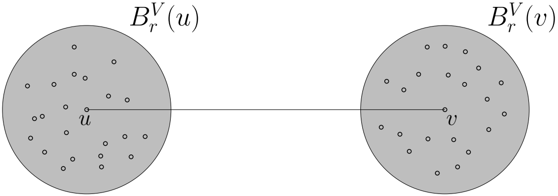
By the pigeonhole principle and the union bound, we have:

$$\mathbb{P}\left[G \text{ has a } uv\text{-clique of size} \geq \mathsf{K}\right]$$
$$\leq \mathbb{P}\left[G|_{\tilde{A}_{uv}} \text{ has a } uv\text{-clique of size} \geq \frac{\mathsf{K}}{2}\right] + \mathbb{P}\left[G|_{B_{uv}} \text{ has a } uv\text{-clique of size} \geq \frac{\mathsf{K}}{2}\right]$$

Case (b)

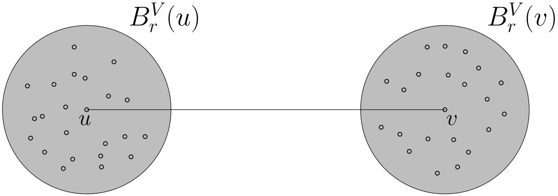


Case (b)



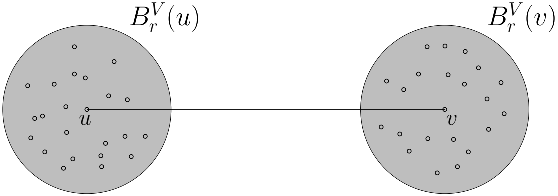
- With **very** high probability, the number of points in any $r-$ball can be bounded from above by applying the Chernoff bound

Case (b)



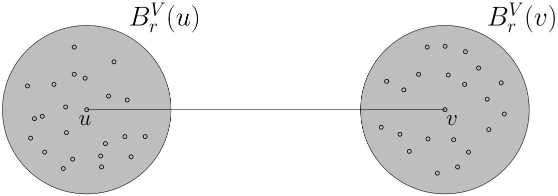
- With **very** high probability, the number of points in any $r-$ball can be bounded from above by applying the Chernoff bound
- Estimate the expectation of the number of $uv-$cliques with size K/2

Case (b)



- With **very** high probability, the number of points in any $r-$ball can be bounded from above by applying the Chernoff bound
- Estimate the expectation of the number of $uv-$cliques with size K/2
- Apply Markov's inequality (first moment method):
  $\mathbb{P}$[has a K/2 uv-clique] $\leq \mathbb{E}$[# of K/2 uv-cliques]

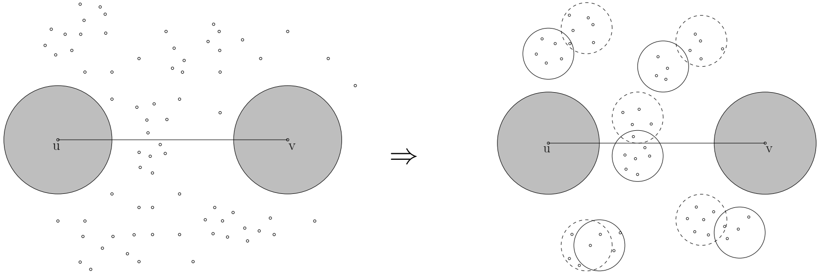
Case (b)



- With **very** high probability, the number of points in any $r-$ball can be bounded from above by applying the Chernoff bound
- Estimate the expectation of the number of $uv-$cliques with size K/2
- Apply Markov's inequality (first moment method):
  $\mathbb{P}$[has a K/2 uv-clique] $\leq \mathbb{E}$[$\#$ of K/2 uv-cliques]
- To let the probability go to 0, we derive some requirement on q and K

Case (a)

- Decouple the randomness by checking **a constant number of** induced subgraphs (well-separated clique-partitions)



Well-separated Clique-partitions Family

Consider an arbitrary RGG $G^* = G^*_{\mathcal{X}}(r)$. A family $\mathcal{P} = \{P_i\}_{i \in \Lambda}$, where $P_i \subseteq V$ and Λ is the index set of P_is, forms a well-separated clique-partitions family of G^* if:

Well-separated Clique-partitions Family

Consider an arbitrary RGG $G^* = G^*_{\mathcal{X}}(r)$. A family $\mathcal{P} = \{P_i\}_{i \in \Lambda}$, where $P_i \subseteq V$ and Λ is the index set of P_is, forms a well-separated clique-partitions family of G^* if:

1. $V = \cup_{i \in \Lambda} P_i$.

Well-separated Clique-partitions Family

Consider an arbitrary RGG $G^* = G^*_{\mathcal{X}}(r)$. A family $\mathcal{P} = \{P_i\}_{i \in \Lambda}$, where $P_i \subseteq V$ and Λ is the index set of P_is, forms a well-separated clique-partitions family of G^* if:

1. $V = \cup_{i \in \Lambda} P_i$.
2. $\forall i \in \Lambda$, P_i can be partitioned as $P_i = C_1^{(i)} \sqcup C_2^{(i)} \sqcup \cdots \sqcup C_{m_i}^{(i)}$ where

Well-separated Clique-partitions Family

Consider an arbitrary RGG $G^* = G^*_{\mathcal{X}}(r)$. A family $\mathcal{P} = \{P_i\}_{i \in \Lambda}$, where $P_i \subseteq V$ and Λ is the index set of P_is, forms a well-separated clique-partitions family of G^* if:

1. $V = \cup_{i \in \Lambda} P_i$.
2. $\forall i \in \Lambda$, P_i can be partitioned as $P_i = C_1^{(i)} \sqcup C_2^{(i)} \sqcup \cdots \sqcup C_{m_i}^{(i)}$ where
   (2-a) $\forall j \in [1, m_i]$, there exist $\bar{v}_j^{(i)} \in V$ such that $C_j^{(i)} \subseteq B_{r/2}\left(\bar{v}_j^{(i)}\right) \cap V$.

Well-separated Clique-partitions Family

Consider an arbitrary RGG $G^* = G^*_{\mathcal{X}}(r)$. A family $\mathcal{P} = \{P_i\}_{i\in\Lambda}$, where $P_i \subseteq V$ and Λ is the index set of P_is, forms a well-separated clique-partitions family of G^* if:

1. $V = \cup_{i\in\Lambda} P_i$.
2. $\forall i \in \Lambda$, P_i can be partitioned as $P_i = C_1^{(i)} \sqcup C_2^{(i)} \sqcup \cdots \sqcup C_{m_i}^{(i)}$ where
   - (2-a) $\forall j \in [1, m_i]$, there exist $\bar{v}_j^{(i)} \in V$ such that $C_j^{(i)} \subseteq B_{r/2}\left(\bar{v}_j^{(i)}\right) \cap V$.
   - (2-b) For any $j_1, j_2 \in [1, m_i]$ with $j_1 \neq j_2$, $d_H\left(C_{j_1}^{(i)}, C_{j_2}^{(i)}\right) > r$, where d_H is the Hausdorff distance between two sets in metric space (X, d).

Well-separated Clique-partitions Family

Consider an arbitrary RGG $G^* = G^*_{\mathcal{X}}(r)$. A family $\mathcal{P} = \{P_i\}_{i\in\Lambda}$, where $P_i \subseteq V$ and Λ is the index set of P_is, forms a well-separated clique-partitions family of G^* if:

1. $V = \cup_{i\in\Lambda} P_i$.
2. $\forall i \in \Lambda$, P_i can be partitioned as $P_i = C_1^{(i)} \sqcup C_2^{(i)} \sqcup \cdots \sqcup C_{m_i}^{(i)}$ where
   - (2-a) $\forall j \in [1, m_i]$, there exist $\bar{v}_j^{(i)} \in V$ such that $C_j^{(i)} \subseteq B_{r/2}\left(\bar{v}_j^{(i)}\right) \cap V$.
   - (2-b) For any $j_1, j_2 \in [1, m_i]$ with $j_1 \neq j_2$, $d_H\left(C_{j_1}^{(i)}, C_{j_2}^{(i)}\right) > r$, where d_H is the Hausdorff distance between two sets in metric space (X, d).

We also call $C_1^{(i)} \sqcup C_2^{(i)} \sqcup \cdots \sqcup C_{m_i}^{(i)}$ a *clique-partition of* P_i *(w.r.t.* G^**)*, and its *size* (cardinality) is m_i. The *size* of the well-separated clique-partitions family $\mathcal{P}$ is its cardinality $|\mathcal{P}| = |\Lambda|$.

Well-separated Clique-partitions Family (cont'd)

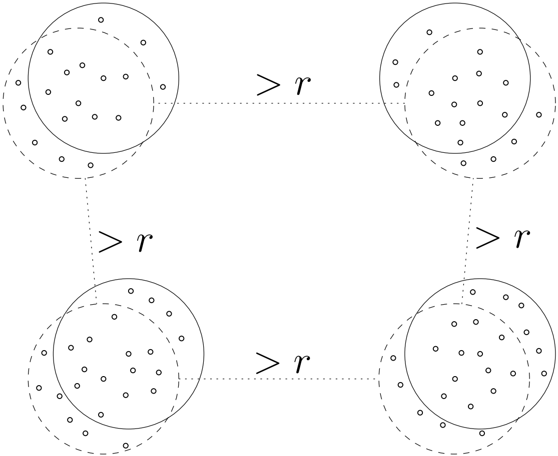


Figure: Points in the solid balls are P_1, and those in dashed balls are P_2. Each adapts a clique-partition of size $m_1 = m_2 = 4$. Assuming that all nodes in G^* are shown in this figure, then $\mathcal{P} = \{P_1, P_2\}$ forms a well-separated clique-partitions family of G^*.

Well-separated Clique-partitions Family (cont'd)

Theorem (Besicovitch Covering Lemma, doubling space version)

[a] *Let $\mathcal{X} = (X, d)$ be a doubling space. Then, there exists a constant $\beta = \beta(\mathcal{X}) \in \mathbb{N}$ such that for any $P \subset X$ and $\delta > 0$, there are β number of δ-packings w.r.t. P, denoted by $\{\mathcal{B}_1, \cdots, \mathcal{B}_\beta\}$, whose union also covers P.*

---

[a] Antti Kaenmaki, Tapio Rajala, and Ville Suomala. Local homogeneity and dimensions of measures. ANNALI DELLA SCUOLA NORMALE SUPERIORE DI PISA-CLASSE DI SCIENZE, 16(4):1315 1351, 2016.

Well-separated Clique-partitions Family (cont'd)

Theorem (Besicovitch Covering Lemma, doubling space version)

[a] *Let $\mathcal{X} = (X, d)$ be a doubling space. Then, there exists a constant $\beta = \beta(\mathcal{X}) \in \mathbb{N}$ such that for any $P \subset X$ and $\delta > 0$, there are β number of δ-packings w.r.t. P, denoted by $\{\mathcal{B}_1, \cdots, \mathcal{B}_\beta\}$, whose union also covers P.*

[a] Antti Kaenmaki, Tapio Rajala, and Ville Suomala. Local homogeneity and dimensions of measures. ANNALI DELLA SCUOLA NORMALE SUPERIORE DI PISA-CLASSE DI SCIENZE, 16(4):1315 1351, 2016.

We call the constant $\beta(\mathcal{X})$ above the *Besicovitch constant*.

Well-separated Clique-partitions Family (cont'd)

Theorem (Besicovitch Covering Lemma, doubling space version)

[a] *Let $\mathcal{X} = (X, d)$ be a doubling space. Then, there exists a constant $\beta = \beta(\mathcal{X}) \in \mathbb{N}$ such that for any $P \subset X$ and $\delta > 0$, there are β number of δ-packings w.r.t. P, denoted by $\{\mathcal{B}_1, \cdots, \mathcal{B}_\beta\}$, whose union also covers P.*

[a] Antti Kaenmaki, Tapio Rajala, and Ville Suomala. Local homogeneity and dimensions of measures. ANNALI DELLA SCUOLA NORMALE SUPERIORE DI PISA-CLASSE DI SCIENZE, 16(4):1315 1351, 2016.

We call the constant $\beta(\mathcal{X})$ above the *Besicovitch constant*.

Theorem (Existence of finite-size W.S.C.P family)

Let $G^* = G^*_{\mathcal{X}}(r)$ be an n-node random geometric graph generated from $(\mathcal{X}, \mu, r)$ where $\mathcal{X} = (X, d)$ and μ is a doubling measure supported on X. There is a well-separated clique-partitions family $\mathcal{P} = \{P_i\}_{i \in \Lambda}$ of G^* with $|\Lambda| \leq \beta^2$, where $\beta = \beta(\mathcal{X})$ is the Besicovitch constant of $\mathcal{X}$.

Case (a) (cont'd)

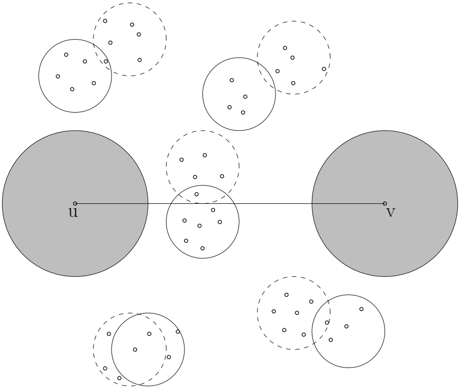
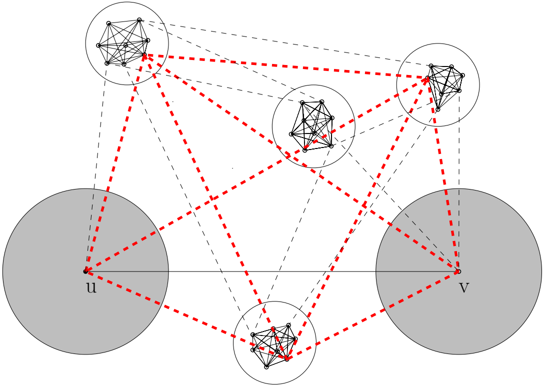


Figure: A well-separated clique partition $\mathcal{P} = \{P_1, P_2\}$ of A_{uv} — points in the solid ball are P_1, and those in dashed ball are P_2.

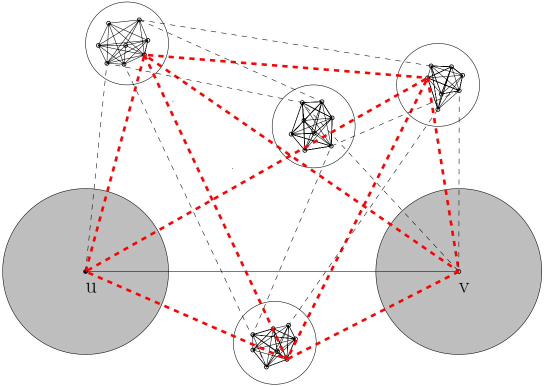
Case (a) (cont'd)

- In each induced subgraph (well-separated clique-partition)



Case (a) (cont'd)

- In each induced subgraph (well-separated clique-partition)



  - Use the same strategy as in case (b)

Clique Number in Standard RGG + $q-$Insertion

Theorem (By-product)

Suppose $n^{-\epsilon} \ll nr_n^d \ll \log n$ *for all* $\epsilon > 0$. *Then, for the* $q-$*perturbed random geometric graph* $\tilde{G}_q(\boldsymbol{X}_n; r_n)$, *the following holds*

- *If* $q \leq C_1 \left(\frac{nr_n^d}{\log n}\right)^{C_2}$, *where* C_1, C_2 *are two constants, then with high probability, we have*

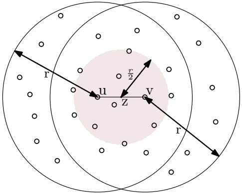
$$\omega\left(\tilde{G}_q(\boldsymbol{X}_n; r_n)\right) = \Theta\left(\frac{\log n}{\log \frac{\log n}{nr_n^d}}\right)$$

- *If* $\left(\frac{nr_n^d}{\log n}\right)^{\xi} \ll q \leq C_3$ *for all* $\xi > 0$ *where* C_3 *is a constant, then with high probability, we have*

$$\omega\left(\tilde{G}_q(\boldsymbol{X}_n; r_n)\right) = \Theta\left(\log_{\frac{1}{q}} n\right)$$

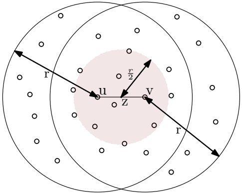
$(p, q)-$perturbation Case

- The lower bound for good edges



(p,q)−perturbation Case

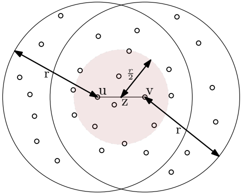
- The lower bound for good edges



  - $B_r(u) \cap B_r(v)$ must contain an $r/2$ centered at the midpoint z of a geodesic connecting u to v in M

$(p, q)-$perturbation Case

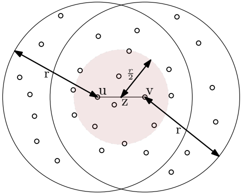
- The lower bound for good edges



  - $B_r(u) \cap B_r(v)$ must contain an $r/2$ centered at the midpoint z of a geodesic connecting u to v in M
  - At most $\binom{n}{2}$ such $r/2-$balls

$(p, q)-$perturbation Case

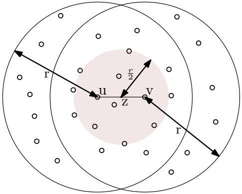
- The lower bound for good edges



  - $B_r(u) \cap B_r(v)$ must contain an $r/2$ centered at the midpoint z of a geodesic connecting u to v in M
  - At most $\binom{n}{2}$ such $r/2-$balls
  - The number of points in any $r/2-$ball can be bounded from below, say N_z, with **very** high probability by applying the Chernoff bound.

$(p, q)-$perturbation Case

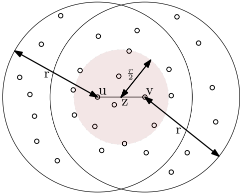
- The lower bound for good edges



  - $B_r(u) \cap B_r(v)$ must contain an $r/2$ centered at the midpoint z of a geodesic connecting u to v in M
  - At most $\binom{n}{2}$ such $r/2-$balls
  - The number of points in any $r/2-$ball can be bounded from below, say N_z, with **very** high probability by applying the Chernoff bound.
  - Roughly speaking, we have a corresponding $G(N_z, 1-p)$ for each good edge locally with $N_z = O(\log n)$

$(p, q)-$perturbation Case

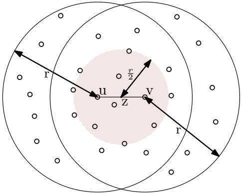
- The lower bound for good edges



  - $B_r(u) \cap B_r(v)$ must contain an $r/2$ centered at the midpoint z of a geodesic connecting u to v in M
  - At most $\binom{n}{2}$ such $r/2-$balls
  - The number of points in any $r/2-$ball can be bounded from below, say N_z, with **very** high probability by applying the Chernoff bound.
  - Roughly speaking, we have a corresponding $G(N_z, 1-p)$ for each good edge locally with $N_z = O(\log n)$
  - Apply **Janson's Inequality** to get a lower bound for the clique number

$(p, q)-$perturbation Case

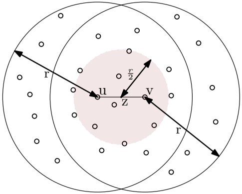
- The lower bound for good edges



  - $B_r(u) \cap B_r(v)$ must contain an $r/2$ centered at the midpoint z of a geodesic connecting u to v in M
  - At most $\binom{n}{2}$ such $r/2-$balls
  - The number of points in any $r/2-$ball can be bounded from below, say N_z, with **very** high probability by applying the Chernoff bound.
  - Roughly speaking, we have a corresponding $G(N_z, 1-p)$ for each good edge locally with $N_z = O(\log n)$
  - Apply **Janson's Inequality** to get a lower bound for the clique number
  - Union bound

$(p, q)-$perturbation Case

- The lower bound for good edges



  - $B_r(u) \cap B_r(v)$ must contain an $r/2$ centered at the midpoint z of a geodesic connecting u to v in M
  - At most $\binom{n}{2}$ such $r/2-$balls
  - The number of points in any $r/2-$ball can be bounded from below, say N_z, with **very** high probability by applying the Chernoff bound.
  - Roughly speaking, we have a corresponding $G(N_z, 1-p)$ for each good edge locally with $N_z = O(\log n)$
  - Apply **Janson's Inequality** to get a lower bound for the clique number
  - Union bound
- The upper bound for bad edges (Same strategy!)

Clique Number in Standard RGG + $p-$Deletion

- For regime "$nr_n^d \leq n^{-\alpha}$ for some α"
  - Directly apply the Poisson approximation (the Stein-Chen method)
- For regime "$n^{-\epsilon} \ll nr_n^d \ll \log n$ for all $\epsilon > 0$"
  - Need put some constraint on p to fit the Poisson approximation setting
- For other regime? (e.g. $\frac{\sigma nr_n^d}{\log n} \to t \in (0, \infty)$)

Remarks

- This result can be used to filter bad edges in the observed graph G.

Remarks

- This result can be used to filter bad edges in the observed graph G.

τ-Clique filtering

Given graph G, we construct another graph $\widehat{G}_\tau$ on the same vertex set as follows: For each edge $(u, v) \in E(G)$, we insert the edge (u, v) into $E(\widehat{G}_\tau)$ if and only if $\omega_{u,v}(G) \geq \tau$. That is, $V(\widehat{G}_\tau) = V(G)$ and $E(\widehat{G}_\tau) := \{(u, v) \in E(G) \mid \omega_{u,v}(G) \geq \tau\}$.

Remarks

- This result can be used to filter bad edges in the observed graph G.

τ-Clique filtering

Given graph G, we construct another graph $\widehat{G}_\tau$ on the same vertex set as follows: For each edge $(u, v) \in E(G)$, we insert the edge (u, v) into $E(\widehat{G}_\tau)$ if and only if $\omega_{u,v}(G) \geq \tau$. That is, $V(\widehat{G}_\tau) = V(G)$ and $E(\widehat{G}_\tau) := \{(u, v) \in E(G) \mid \omega_{u,v}(G) \geq \tau\}$.

- By carefully choosing τ, the shortest-path metric $d_{\widehat{G}_\tau}$ is a 3-approximation of d_{G^*}.

Remarks

- This result can be used to filter bad edges in the observed graph G.

τ-Clique filtering

Given graph G, we construct another graph $\widehat{G}_\tau$ on the same vertex set as follows: For each edge $(u, v) \in E(G)$, we insert the edge (u, v) into $E(\widehat{G}_\tau)$ if and only if $\omega_{u,v}(G) \geq \tau$. That is, $V(\widehat{G}_\tau) = V(G)$ and $E(\widehat{G}_\tau) := \{(u, v) \in E(G) \mid \omega_{u,v}(G) \geq \tau\}$.

- By carefully choosing τ, the shortest-path metric $d_{\widehat{G}_\tau}$ is a 3-approximation of d_{G^*}.
- Significantly larger range of insertion probability q than the case with Jaccard-filtering[4].

[4]Srinivasan Parthasarathy, David Sivakoff, Minghao Tian, and Yusu Wang. A quest to unravel the metric structure behind perturbed networks. In 33rd International Symposium on Computational Geometry, SoCG 2017, July 4-7, 2017, Brisbane, Australia, pages 53:153:16, 2017.

Remarks

- This result can be used to filter bad edges in the observed graph G.

τ-Clique filtering

Given graph G, we construct another graph $\widehat{G}_\tau$ on the same vertex set as follows: For each edge $(u, v) \in E(G)$, we insert the edge (u, v) into $E(\widehat{G}_\tau)$ if and only if $\omega_{u,v}(G) \geq \tau$. That is, $V(\widehat{G}_\tau) = V(G)$ and $E(\widehat{G}_\tau) := \{(u, v) \in E(G) \mid \omega_{u,v}(G) \geq \tau\}$.

- By carefully choosing τ, the shortest-path metric $d_{\widehat{G}_\tau}$ is a 3-approximation of d_{G^*}.
- Significantly larger range of insertion probability q than the case with Jaccard-filtering[4].
- However, Jaccard-filtering is computationally much more feasible.

[4] Srinivasan Parthasarathy, David Sivakoff, Minghao Tian, and Yusu Wang. A quest to unravel the metric structure behind perturbed networks. In 33rd International Symposium on Computational Geometry, SoCG 2017, July 4-7, 2017, Brisbane, Australia, pages 53:153:16, 2017.

Open Questions

- Other regimes? (e.g. sparse, thermodynamic limit, etc.)
- Other quantities to look at? (e.g. chromatic number, Lovász number)
- Other metric structures? (e.g. diffusion distance)

Thank you for your attention!