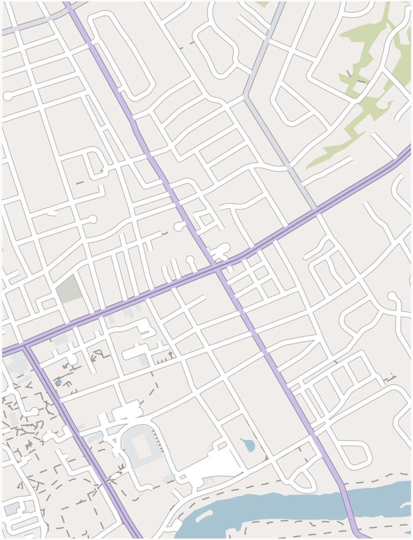
ON CONSTANT-SIZE GRAPHS THAT PRESERVE THE LOCAL STRUCTURE OF HIGH-GIRTH GRAPHS

H. F., Pan Peng, Christian Sohler

Hendrik Fichtenberger

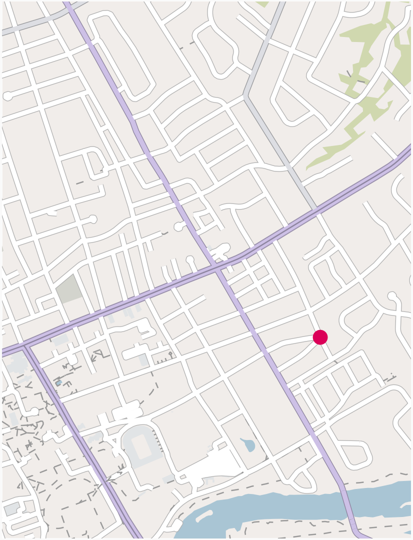
26. August 2015

TU Dortmund, Germany



- There are many ways to define the *local structure* of a graph

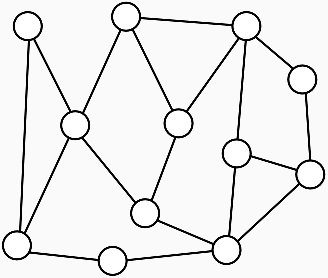
Map: © OpenStreetMap contributors, see http://www.openstreetmap.org/copyright



- There are many ways to define the *local structure* of a graph
- We are interested in the local structure of vertices ● …

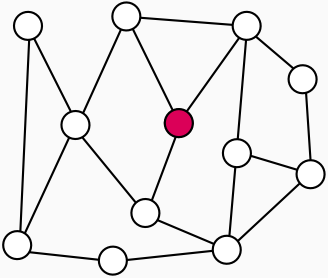
- There are many ways to define the *local structure* of a graph
- We are interested in the local structure of vertices ● …
- … and the subgraph ● around them

- There are many ways to define the *local structure* of a graph
- We are interested in the local structure of vertices ● …
- … and the subgraph ● around them
- We will summarize over all vertices in a graph



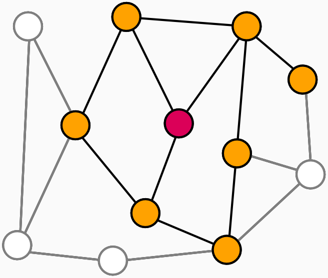
$k = 2$

- Let
  - $G = (V, E)$ be a graph
  - $k \geq 0$ be an integer



$k = 2$

- Let
  - $G = (V, E)$ be a graph
  - $k \geq 0$ be an integer
  - $u \in V$ ● be a node



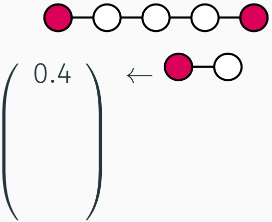
$k = 2$

- Let
  - $G = (V, E)$ be a graph
  - $k \geq 0$ be an integer
  - $u \in V$ ● be a node
- k-disc of u:
  - Subgraph induced by all vertices ● within distance at most k to u
  - Rooted at u

$k = 1$

- Let
  - $G = (V, E)$ be a graph
  - $k \geq 0$ be an integer

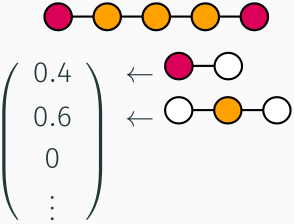
LOCAL STRUCTURE IN GRAPHS



$k = 1$

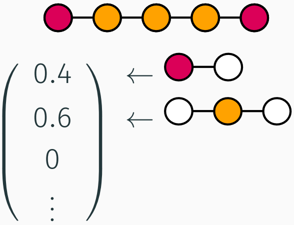
- Let
  - $G = (V, E)$ be a graph
  - $k \geq 0$ be an integer
- $\text{freq}_k(G)$: k-disc frequency vector
  - Vector indexed by all k-disc isomorphism types
  - Counts the fraction of each type of k-disc in G

LOCAL STRUCTURE IN GRAPHS



$k = 1$

- Let
  - $G = (V, E)$ be a graph
  - $k \geq 0$ be an integer
- $\text{freq}_k(G)$: k-disc frequency vector
  - Vector indexed by all k-disc isomorphism types
  - Counts the fraction of each type of k-disc in G



$k = 1$

- Let
  - $G = (V, E)$ be a graph
  - $k \geq 0$ be an integer
- $\text{freq}_k(G)$: k-disc frequency vector
  - Vector indexed by all k-disc isomorphism types
  - Counts the fraction of each type of k-disc in G

From now on

- All graphs are undirected and have maximum degree d
- d and k are some constants

Question [http://sublinear.info/42]

Given $\epsilon, k > 0$ and a bounded-degree graph G, is there always a small graph H of size $f(\epsilon, d, k)$ such that

$$\|\text{freq}_k(G) - \text{freq}_k(H)\|_1 \leq \epsilon\,?$$

Intuition Approximate the local structure of a large bounded-degree graph by a constant-size graph

Question [http://sublinear.info/42]

Given $\epsilon, k > 0$ and a bounded-degree graph G, is there always a small graph H of size $f(\epsilon, d, k)$ such that

$$\|\text{freq}_k(G) - \text{freq}_k(H)\|_1 \leq \epsilon\,?$$

Intuition Approximate the local structure of a large bounded-degree graph by a constant-size graph

Answer Yes! There is a simple proof by Alon[1].
However, no (explicit) bound on $|V(H)|$ is known.

[1] see Lovász, *Large Networks and Graph Limits*, 2012

Question [http://sublinear.info/42]

Given $\epsilon, k > 0$ and a bounded-degree graph G, is there always a small graph H of size $f(\epsilon, d, k)$ such that

$$\|\text{freq}_k(G) - \text{freq}_k(H)\|_1 \leq \epsilon\,?$$

Intuition Approximate the local structure of a large bounded-degree graph by a constant-size graph

Answer Yes! There is a simple proof by Alon[1].
However, no (explicit) bound on $|V(H)|$ is known.

In this talk Bound for special case where all k-discs are trees

[1] see Lovász, *Large Networks and Graph Limits*, 2012

Theorem

Given query access to the adjacency lists of a graph G with girth $> 2k+1$ and with maximum degree d, the algorithm outputs a graph H of size at most $f_1(d,k) \cdot \epsilon^{-2}\delta^{-1}$ that satisfies

$$\|\mathrm{freq}_k(G) - \mathrm{freq}_k(H)\|_1 \leq \epsilon$$

with probability $1-\delta$. Its running time is independent of G.

Theorem

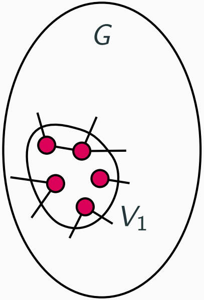
Given query access to the adjacency lists of a graph G with girth $> 2k+1$ and with maximum degree d, the algorithm outputs a graph H of size at most $f_1(d,k) \cdot \epsilon^{-2}\delta^{-1}$ that satisfies

$$\|\mathrm{freq}_k(G) - \mathrm{freq}_k(H)\|_1 \leq \epsilon$$

with probability $1-\delta$. Its running time is independent of G.

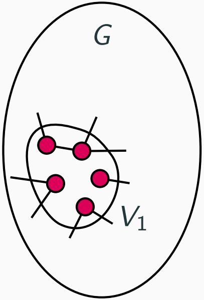
Relations
- Regularity lemma and graph limits[2]
- Property Testing[3]

[2] see Elek, *On the Limit of Large Girth Graph Sequences*, 2010
[3] e.g., Newman, Sohler, *Every Property of Hyperfinite Graphs Is Testable*, 2011



Task Given $G = (V, E)$, construct small graph H with similar k-disc distribution

Idea 1. Sample a small set of vertices V_1 ●

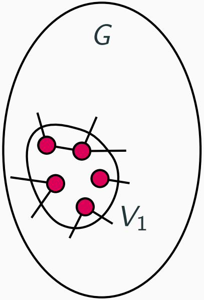


Task Given $G = (V, E)$, construct small graph H with similar k-disc distribution

Idea
1. Sample a small set of vertices V_1 ●
2. For every k-disc type Δ, we picked

$$(\text{fraction of } \Delta \text{ in } G \pm \epsilon) \cdot |V_1|$$

vertices with k-disc Δ w.h.p.

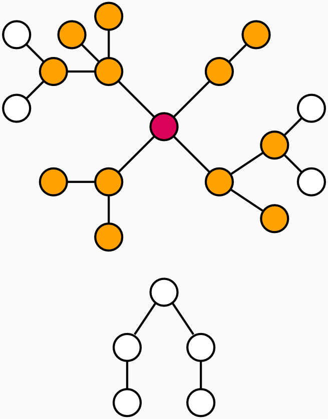


Task Given $G = (V, E)$, construct small graph H with similar k-disc distribution

Idea
1. Sample a small set of vertices V_1
2. For every k-disc type Δ, we picked

$$(\text{fraction of } \Delta \text{ in } G \pm \epsilon) \cdot |V_1|$$

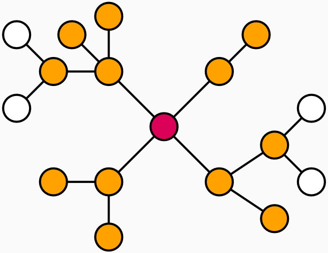
   vertices with k-disc Δ w.h.p.
3. We would like to choose $H := G[V_1]$
   - $E(V_1, V \setminus V_1)$ might be large
   - Deleting all these edges alters many k-discs
   - Need a way to reduce size of cut…

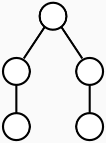


$k = 2$

Observation 1

A k-disc is the union of the $(k-1)$-discs of its root's neighbors





$k = 2$

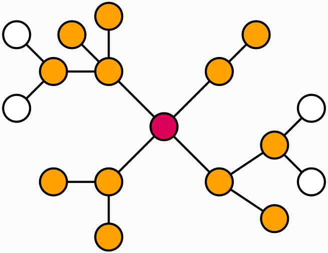
Observation 1

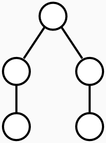
A k-disc is the union of the $(k-1)$-discs of its root's neighbors

Observation 2

If G has girth $> 2k+1$, then all its k-discs are trees

LOCAL STRUCTURE AND HIGH GIRTH





$k = 2$

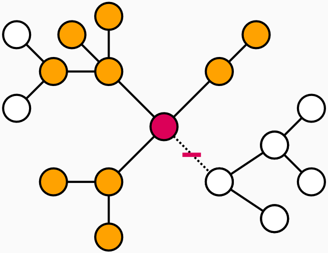
Observation 1

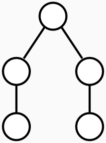
A k-disc is the union of the $(k-1)$-discs of its root's neighbors

Observation 2

If G has girth $> 2k+1$, then all its k-discs are trees

Rewiring Change edges without changing k-disc of ●





$k = 2$

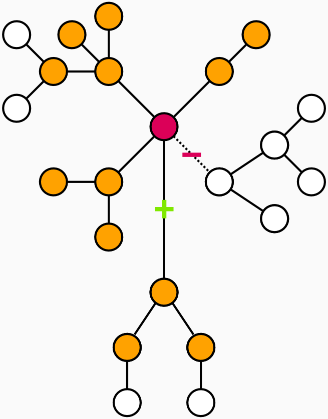
Observation 1

A k-disc is the union of the $(k-1)$-discs of its root's neighbors

Observation 2

If G has girth $> 2k+1$, then all its k-discs are trees

Rewiring Change edges without changing k-disc of ●



$k = 2$

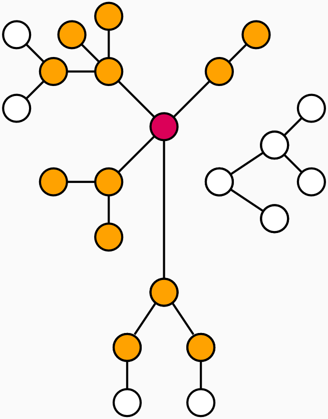
Observation 1

A k-disc is the union of the $(k-1)$-discs of its root's neighbors

Observation 2

If G has girth $> 2k+1$, then all its k-discs are trees

Rewiring Change edges without changing k-disc of ●



$k = 2$

Observation 1

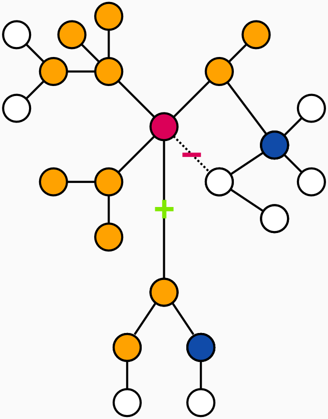
A k-disc is the union of the $(k-1)$-discs of its root's neighbors

Observation 2

If G has girth $> 2k+1$, then all its k-discs are trees

Rewiring Change edges without changing k-disc of ●

LOCAL STRUCTURE AND HIGH GIRTH



$k = 2$

Observation 1

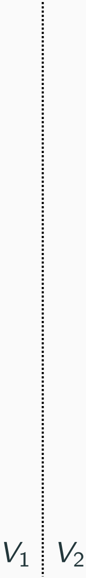
A k-disc is the union of the $(k-1)$-discs of its root's neighbors

Observation 2

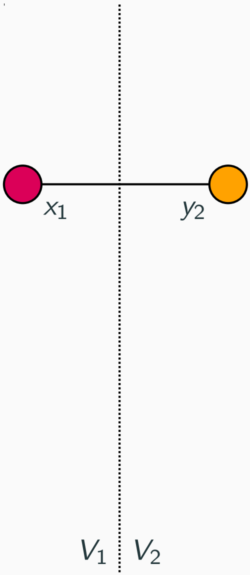
If G has girth $> 2k+1$, then all its k-discs are trees

Rewiring Change edges without changing k-disc of ●

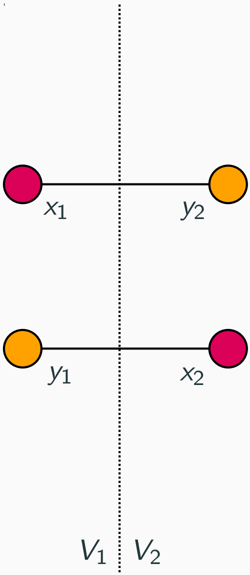
Remark To do this, k-disc must be cycle-free



1. Let
   - V_1 be our sample, $V_2 := V \setminus V_1$
   - Δ_1, Δ_2 be k-disc isomorphism types

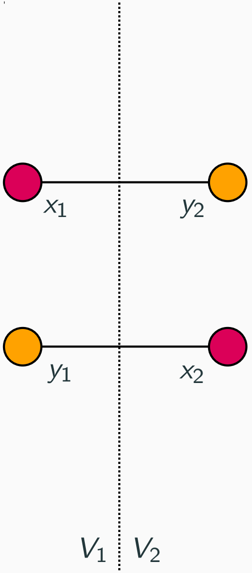


1. Let
   - V_1 be our sample, $V_2 := V \setminus V_1$
   - Δ_1, Δ_2 be k-disc isomorphism types
2. Assume that there is
   - an edge $(x_1, y_2) \in V_1 \times V_2$ s.t. $\text{disc}_k(x_1) \simeq \Delta_1$ ●, $\text{disc}_k(y_2) \simeq \Delta_2$ ●



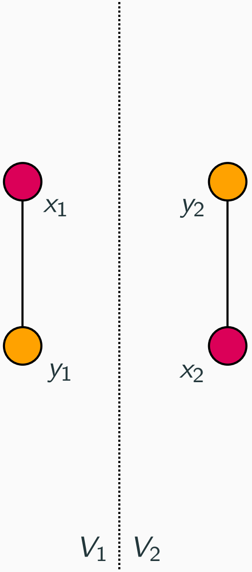
1. Let
   - V_1 be our sample, $V_2 := V \setminus V_1$
   - Δ_1, Δ_2 be k-disc isomorphism types
2. Assume that there is
   - an edge $(x_1, y_2) \in V_1 \times V_2$ s.t. $\text{disc}_k(x_1) \simeq \Delta_1$ (red), $\text{disc}_k(y_2) \simeq \Delta_2$ (orange)
   - an edge $(y_1, x_2) \in V_1 \times V_2$ s.t. $\text{disc}_k(y_1) \simeq \Delta_2$ (orange), $\text{disc}_k(x_2) \simeq \Delta_1$ (red)

REWIRING EDGES IN THE CUT



1. Let
   - V_1 be our sample, $V_2 := V \setminus V_1$
   - Δ_1, Δ_2 be k-disc isomorphism types
2. Assume that there is
   - an edge $(x_1, y_2) \in V_1 \times V_2$ s.t. $\text{disc}_k(x_1) \simeq \Delta_1$ ●, $\text{disc}_k(y_2) \simeq \Delta_2$ ●
   - an edge $(y_1, x_2) \in V_1 \times V_2$ s.t. $\text{disc}_k(y_1) \simeq \Delta_2$ ●, $\text{disc}_k(x_2) \simeq \Delta_1$ ●
   - $\{\text{disc}_k(x_1), \text{disc}_k(y_2)\}$ and $\{\text{disc}_k(y_1), \text{disc}_k(x_2)\}$ do not overlap

REWIRING EDGES IN THE CUT



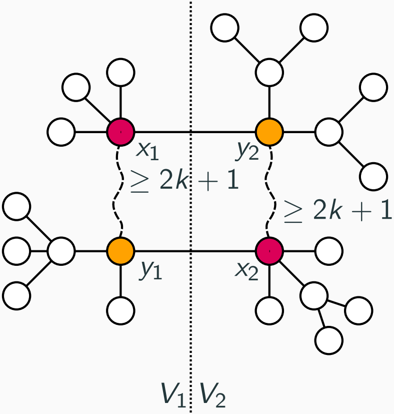
1. Let
    - V_1 be our sample, $V_2 := V \setminus V_1$
    - Δ_1, Δ_2 be k-disc isomorphism types
2. Assume that there is
    - an edge $(x_1, y_2) \in V_1 \times V_2$ s.t. $\text{disc}_k(x_1) \simeq \Delta_1$ ●, $\text{disc}_k(y_2) \simeq \Delta_2$ ●
    - an edge $(y_1, x_2) \in V_1 \times V_2$ s.t. $\text{disc}_k(y_1) \simeq \Delta_2$ ●, $\text{disc}_k(x_2) \simeq \Delta_1$ ●
    - $\{\text{disc}_k(x_1), \text{disc}_k(y_2)\}$ and $\{\text{disc}_k(y_1), \text{disc}_k(x_2)\}$ do not overlap
3. Then, we
    - remove $(x_1, y_2), (y_1, x_2)$ and then
    - insert $(x_1, y_1), (y_2, x_2)$

Lemma

One can rewire edges without changing the k-disc distribution of G until the cut between V_1 and V_2 has size at most $f(d, k)$.

Lemma

One can rewire edges without changing the k-disc distribution of G until the cut between V_1 and V_2 has size at most $f(d, k)$.

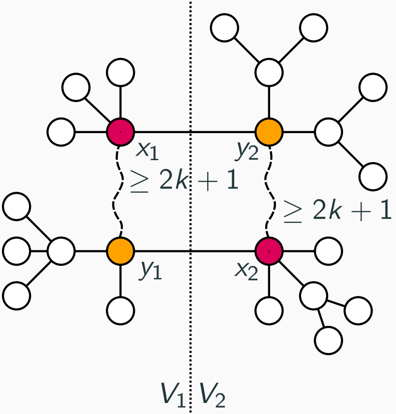


Proof considers two cases:

1. There are two edges with the same pair of k-discs that are not too close
   $\rightarrow$ rewiring changes no k-disc up to isomorphism

Lemma

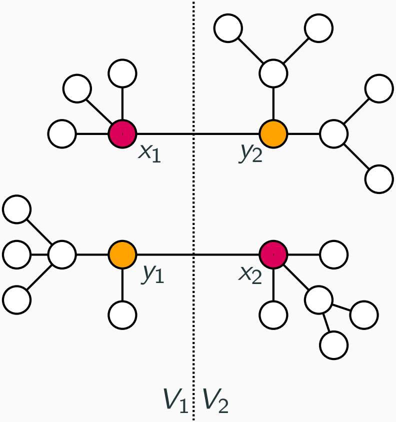
One can rewire edges without changing the k-disc distribution of G until the cut between V_1 and V_2 has size at most $f(d, k)$.



Proof considers two cases:

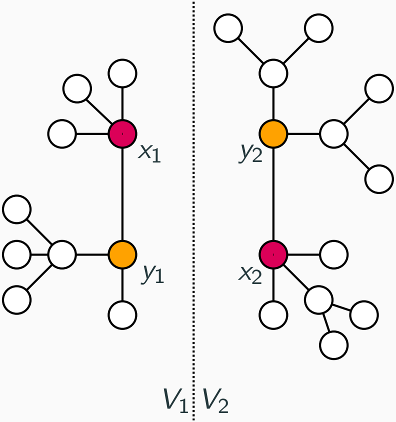
1. There are two edges with the same pair of k-discs that are not too close
   $\rightarrow$ rewiring changes no k-disc up to isomorphism
2. No such edges exist
   $\rightarrow$ $|E(V_1, V_2)|$ is small

Case 1: Edges can be rewired



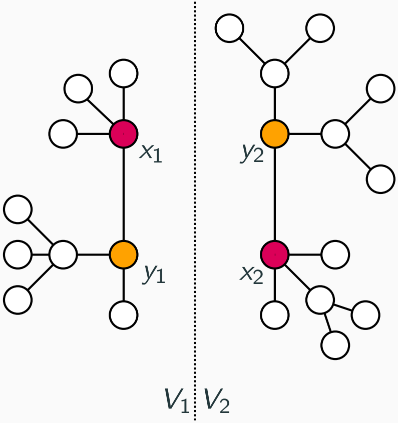
- Find an edge (x_1, y_2) and a counterpart (y_1, x_2) at distance $\geq 2k+1$ for rewiring

Case 1: Edges can be rewired



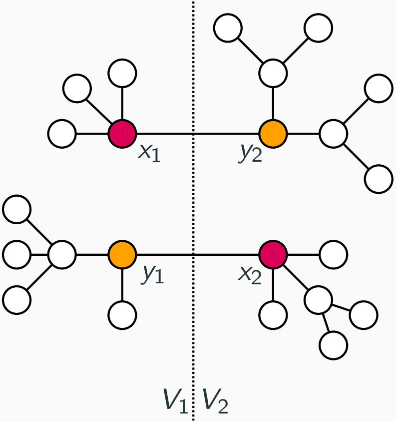
- Find an edge (x_1, y_2) and a counterpart (y_1, x_2) at distance $\geq 2k + 1$ for rewiring
- Remove $(x_1, y_2), (y_1, x_2)$ and add $(x_1, y_1), (x_2, y_2)$ $\rightarrow$ size of cut decreases

Case 1: Edges can be rewired



- Find an edge (x_1, y_2) and a counterpart (y_1, x_2) at distance $\geq 2k+1$ for rewiring
- Remove $(x_1, y_2), (y_1, x_2)$ and add $(x_1, y_1), (x_2, y_2)$ $\rightarrow$ size of cut decreases
- Prove that
  - k-discs before and after rewiring are isomorphic
  - graph has still high girth

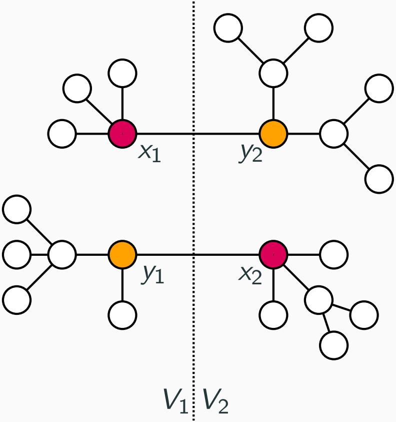
Case 2: No edge can be rewired



- For all k-discs Δ_1, Δ_2:

$$|E(V_1 \times V_2) \cap (\Delta_1 \bullet, \Delta_2 \bullet)|$$
$$\approx |E(V_1 \times V_2) \cap (\Delta_2 \bullet, \Delta_1 \bullet)|$$

Case 2: No edge can be rewired

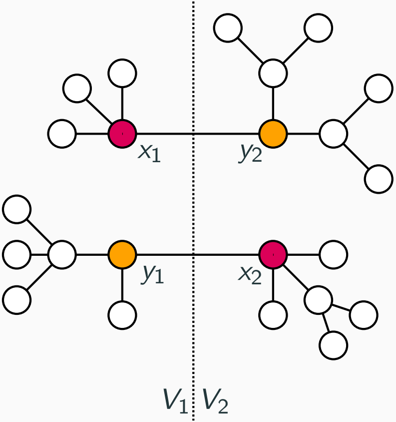


- For all k-discs Δ_1, Δ_2:

$$|E(V_1 \times V_2) \cap (\Delta_1 \bullet, \Delta_2 \bullet)|$$
$$\approx |E(V_1 \times V_2) \cap (\Delta_2 \bullet, \Delta_1 \bullet)|$$

- Let (x_1, y_2) be an edge from $E(V_1, V_2) \cap (\Delta_1, \Delta_2)$

Case 2: No edge can be rewired

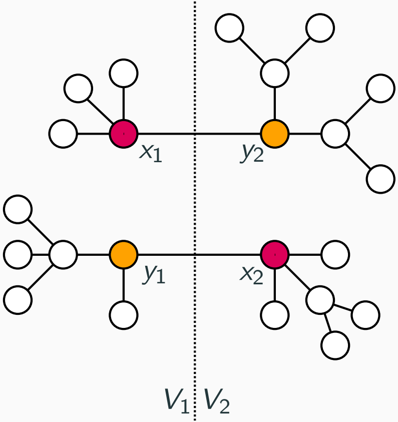


- For all k-discs Δ_1, Δ_2:

$$|E(V_1 \times V_2) \cap (\Delta_1 \bullet, \Delta_2 \bullet)|$$
$$\approx |E(V_1 \times V_2) \cap (\Delta_2 \bullet, \Delta_1 \bullet)|$$

- Let (x_1, y_2) be an edge from $E(V_1, V_2) \cap (\Delta_1, \Delta_2)$
- If we cannot find (y_1, x_2), then $E(V_1, V_2) \cap (\Delta_2, \Delta_1)$ is small

Case 2: No edge can be rewired



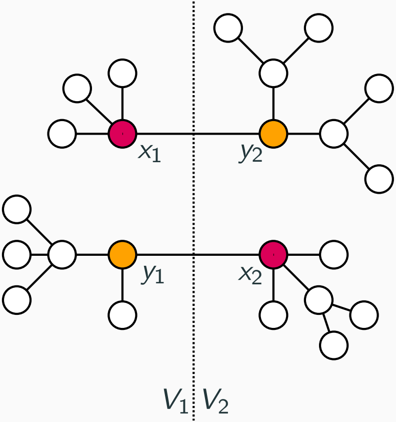
- For all k-discs Δ_1, Δ_2:

$$|E(V_1 \times V_2) \cap (\Delta_1 \bullet, \Delta_2 \bullet)|$$
$$\approx |E(V_1 \times V_2) \cap (\Delta_2 \bullet, \Delta_1 \bullet)|$$

- Let (x_1, y_2) be an edge from $E(V_1, V_2) \cap (\Delta_1, \Delta_2)$
- If we cannot find (y_1, x_2), then $E(V_1, V_2) \cap (\Delta_2, \Delta_1)$ is small
- Hence, $E(V_1, V_2) \cap (\Delta_1, \Delta_2)$ is small

SKETCH OF THE PROOF

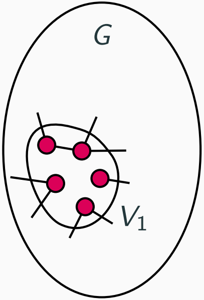
Case 2: No edge can be rewired



- For all k-discs Δ_1, Δ_2:

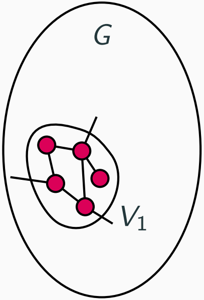
  $|E(V_1 \times V_2) \cap (\Delta_1$ ●, Δ_2 ●$)|$

  $\approx |E(V_1 \times V_2) \cap (\Delta_2$ ●, Δ_1 ●$)|$
- Let (x_1, y_2) be an edge from $E(V_1, V_2) \cap (\Delta_1, \Delta_2)$
- If we cannot find (y_1, x_2), then $E(V_1, V_2) \cap (\Delta_2, \Delta_1)$ is small
- Hence, $E(V_1, V_2) \cap (\Delta_1, \Delta_2)$ is small
- Remove all edges in $E(V_1, V_2)$ $\rightarrow$ only few k-discs are changed



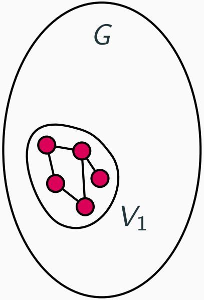
Input Graph $G = (V, E)$ with girth $> 2k + 1$

Algorithm 1. Sample small set of vertices V_1 ●



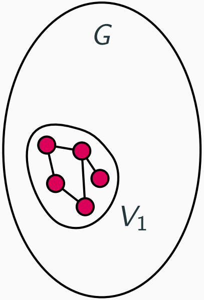
Input Graph $G = (V, E)$ with girth $> 2k + 1$

Algorithm
1. Sample small set of vertices V_1
2. Rewire edges in cut of V_1 and $V_2 := V \setminus V_1$ as long as possible



Input Graph $G = (V, E)$ with girth $> 2k + 1$

Algorithm
1. Sample small set of vertices V_1
2. Rewire edges in cut of V_1 and $V_2 := V \setminus V_1$ as long as possible
3. Remove remaining edges between V_1 and V_2



Input Graph $G = (V, E)$ with girth $> 2k + 1$

Algorithm
1. Sample small set of vertices V_1
2. Rewire edges in cut of V_1 and $V_2 := V \setminus V_1$ as long as possible
3. Remove remaining edges between V_1 and V_2

Output $H := G[V_1]$

Theorem

Given query access to the adjacency lists of a graph G with girth $> 2k + 1$ and with maximum degree d, the algorithm outputs a graph H of size at most $f_1(d, k) \cdot \epsilon^{-2}\delta^{-1}$ that satisfies

$$\|\text{freq}_k(G) - \text{freq}_k(H)\|_1 \leq \epsilon$$

with probability $1 - \delta$. Its running time is independent of G.

Theorem

Given query access to the adjacency lists of a graph G with girth $> 2k + 1$ and with maximum degree d, the algorithm outputs a graph H of size at most $f_1(d, k) \cdot \epsilon^{-2}\delta^{-1}$ that satisfies

$$\|\mathrm{freq}_k(G) - \mathrm{freq}_k(H)\|_1 \leq \epsilon$$

with probability $1 - \delta$. Its running time is independent of G.

Remark Using a deterministic, linear-time algorithm we can improve the bound to $|V(H)| \leq f_2(d, k)/\epsilon$.

Let L be the dimension of the k-disc frequency vector.

Theorem

Given query access to the adjacency lists of a graph G with girth $> 2k+1$ and with maximum degree d, the algorithm outputs a graph H of size at most $\frac{300d^{3k+2}L^3}{\varepsilon^2\delta}$ that satisfies

$$\|\mathrm{freq}_k(G) - \mathrm{freq}_k(H)\|_1 \leq \epsilon$$

with probability $1-\delta$. Its running time is independent of G.

Remark Using a deterministic, linear-time algorithm we can improve the bound to $|V(H)| \leq \frac{36d^{3k+2}L}{\epsilon}$.

LITERATURE

- Bertinoro Workshop on Sublinear Algorithms 2011.
  Open Problems in Data Streams, Property Testing, and Related Topics. `http://sublinear.info/42`.
- L. Lovász.
  Large Networks and Graph Limits.
- G. Elek.
  On the Limit of Large Girth Graph Sequences.
- I. Newman and C. Sohler.
  Every Property of Hyperfinite Graphs Is Testable.