Graphs of Shortest Paths

Simon Dreyer, Antoine Genitrini and Mehdi Naima

[1]Sorbonne Universite, CNRS, LIP6

June 3rd, 2025

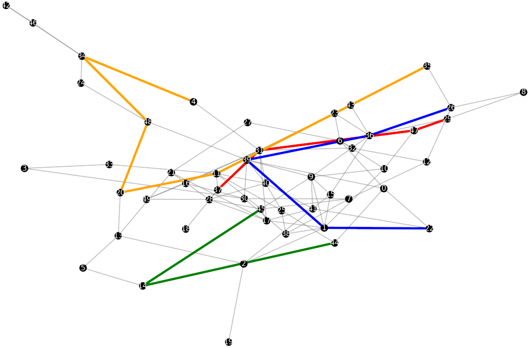
Outline

Section 1. Introduction

Safe places in Graphs

- Find safe places in graphs when pandemic propagates on routes
- Needed a simplistic traffic model

⟶ Assign a shortest path to each agent



What distribution to use? Simplest possible, all shortest paths are equally likely (uniform distribution)

Notations

Given a Graph $G = (V, E)$, denoting $d(s,t)$ the distance from s to t:

- $\mathbf{W}_{st}$ the set of all shortest paths (SP) (all have length $d(s,t)$) from s to t. $\sigma_s(t) = |\mathbf{W}_{st}|$
- $\mathbf{W}_{s\bullet}$ the set of all SP starting from s, $\mathbf{W}_{s\bullet} = \cup_{t \in V} \mathbf{W}_{st}$. $\sigma_{s\bullet} = |\mathbf{W}_{s\bullet}|$
- $\mathbf{W}$ the set of all SP in the graph $\mathbf{W} = \cup_{v \in V} \mathbf{W}_{v\bullet}$. $\sigma = |\mathbf{W}|$

Traffic Assignment

Each agent is assigned a shortest path W, such that $\mathbb{P}(W \in \mathbf{W}) = \frac{1}{\sigma}$

Literature: Sampling Shortest Paths

- Most sampling procedures **fix a source and target nodes**
- Sampling SP is used in:
  - simulating traffic flow [DOW24]
  - studying the topology of a large network [DAHB+06]
  - assessing network damage [CPBV14]
- Two main procedures are mentioned:

Naive

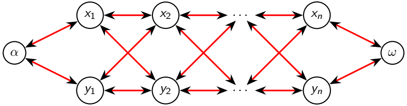
- [DAHB+06; CPBV14; PFV10; ZZW+11; LLFS07; CT11]
- randomly selecting one shortest path from all possible paths

Random Weights

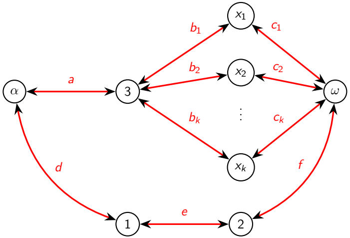
- [LBCX03; CM03; WVM10; FV07]
- edges assigned random weights $(1+\epsilon)$
- return the unique path left
- new sampling: new weights

Literature: Naive and Random Weights

- Problem: There can be an exponential number of paths
- Graph has $2n+2$ nodes
- There are 2^n SP from α and ω



- Family of graphs G_k. weights $\in [1-\frac{1}{n}, 1+\frac{1}{n}]$
- $W_0 = \alpha \to 1 \to 2 \to \omega$
- $W_i = \alpha \to 3 \to x_i \to \omega$
- $\mathbb{P}(W_0) \overset{k=2}{=} \frac{737}{2016} \approx 0.36 \neq \frac{1}{3}$
- $\mathbb{P}(W_0) \overset{k\to+\infty}{\longrightarrow} \frac{1}{24} \neq 0$



Section 2. Contributions

Sampling Algorithm

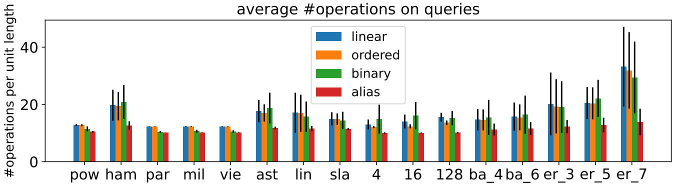
Problem: source-target (s, t) uniform shortest path

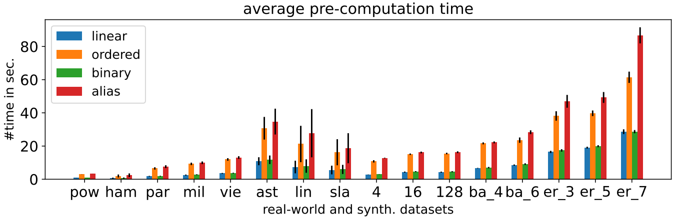
Give a random generation algorithm satisfying $\forall W \in \mathbf{W}_{st}, \mathbb{P}(W) = 1/\sigma_s(t)$ and for all $W \notin \mathbf{W}_{st}, (W) = 0$.

- Two phase Algorithm:
  - Preprocessing: done only once
  - Sampling: any number of times
- Different implementations (linear, ordered, binary, alias)
- Optimal Time Complexity: **Alias**. $O(m)$ for preprocessing and $O(\ell)$ for the sampling. m: number edges, ℓ is the length of the sampled path

Experimental Study

Open Source Implementation in C: https://github.com/simon-dreyer/Shortest_path_sampling





data	dir.	#nodes	#edges	ref.
power_grid	u	4.94K	6.59K	[Kun13]
hamster_full	u	2.43K	16.6K	[Kun13]
paris	d	9.52K	18.3K	[Boe17]
milan	d	12.9K	25.3K	[Boe17]
vienna	d	16.1K	35.7K	[Boe17]
astro_ph	u	18.8K	198K	[Kun13]
linux_mail	d	26.9K	237K	[Kun13]
slashdot	d	51.1K	130K	[Kun13]
4	u	16.4K	28.7K	×
16	u	16.4K	31.7K	×
128	u	16.4K	32.5K	×
ba_4	u	16.4K	65.5K	×
ba_6	u	16.4K	98.3K	×
er_3	u	16.4K	238K	×
er_5	u	16.4K	398K	×
er_7	u	16.4K	558K	×

Query complexity of preprocessing and sampling on the different graphs

Mean Distance vs Mean Length of SP

- Two algorithms to sample shortest paths in a graph:
  1. Iterate the following: Select unif. randomly a pair of nodes (s, t) and sample a unif. shortest path from s to t.
  2. Iterate the following: Select s according to $\sigma_{v\bullet}$ and t according to $\sigma_s(v)$ and sample a unif. shortest path from s to t.

Question

What is the average length of a sampled shortest path?

SORBONNE UNIVERSITÉ LIP6

Mean Distance vs Mean SP Length II

- Nomenclature in the literature seems a little ambiguous: `NetworkX` `average_shortest_path_length(G)`

$\longrightarrow$ It is in fact the average distance

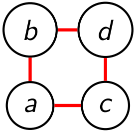
- Average Length of SP does not seem to be studied in literature

Average distance (Algorithm 1):

$$d_G = \frac{1}{n^2} \cdot \sum_{(s,t) \in V^2} d(s,t)$$

Average sh. path length (Algorithm 2):

$$\ell_G = \frac{1}{\sigma} \sum_{(s,t) \in V^2} \sigma_s(t) \cdot d(s,t)$$



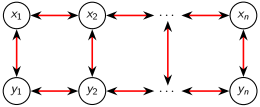
4 paths ($d = 0$), 8 paths ($d = 1$), 8 paths ($d = 2$)

In the 4-cycle graph C_4, $d_{C_4} = 1$ and $\ell_{C_4} = 1.2$

Mean Distance vs Mean SP Length III

Example: 2d Grid $n \times 2$

$$d_G = \frac{1}{n^2} \cdot \sum_{(s,t)\in V^2} d(s,t), \quad \ell_G = \frac{1}{\sigma} \sum_{(s,t)\in V^2} \sigma_s(t) \cdot d(s,t)$$

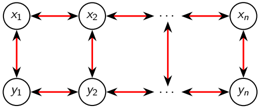


$d_G \overset{n\to\infty}{=} \frac{n}{3}$, $\quad \ell_G \overset{n\to\infty}{=} \frac{n}{2}$, Sampled paths are 3/2 longer using Algorithm 2.

Mean Distance vs Mean SP Length III

Example: 2d Grid $n \times 2$

$$d_G = \frac{1}{n^2} \cdot \sum_{(s,t) \in V^2} d(s,t), \quad \ell_G = \frac{1}{\sigma} \sum_{(s,t) \in V^2} \sigma_s(t) \cdot d(s,t)$$

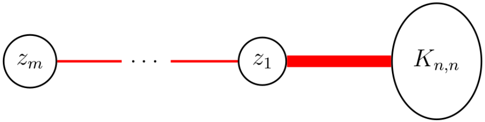


$d_G \overset{n\to\infty}{=} \frac{n}{3}$, $\ell_G \overset{n\to\infty}{=} \frac{n}{2}$, Sampled paths are 3/2 longer using Algorithm 2.

How different can these two measures (d_G and ℓ_G) be?

Arbitrary different in fact, for some graph families we can have $\ell_G / d_G \to \infty$ or $\ell_G / d_G \to 0$

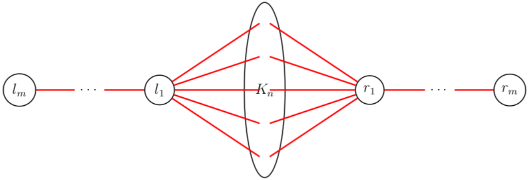
Mean Distance vs Mean SP Length



When $m = n$, $\frac{\ell_G}{d_G} \overset{n\to\infty}{=} 0$

Rate of m	d_G	ℓ_G	$\frac{\ell_G}{d_G}$
$m < \sqrt{n}$	$\frac{5}{4}$	2	$\frac{8}{5}$
$m = \sqrt{n}$	$\frac{3}{2}$	2	$\frac{4}{3}$
$\sqrt{n} < m < n$	$\frac{m^2}{4n}$	2	$\frac{8n}{m^2}$
$m = n$	$\frac{13n}{75}$	$\frac{61}{24}$	$\frac{1525}{104n}$
$n < m < n\sqrt{n}$	$\frac{m}{3}$	$\frac{m^3}{24n^3}$	$\frac{m^2}{8n^3}$
$m = n\sqrt{n}$	$\frac{n\sqrt{n}}{3}$	$\frac{n\sqrt{n}}{27}$	$\frac{1}{9}$
$n\sqrt{n} < m$	$\frac{m}{3}$	$\frac{m}{3}$	1

Mean Distance vs Mean SP Length



When $m = \sqrt{n}$, $\frac{\ell_G}{d_G} \overset{n\to\infty}{=} \infty$

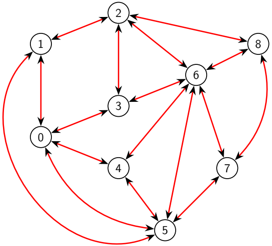
Rate of m	d_G	ℓ_G	$\frac{\ell_G}{d_G}$
$m < \sqrt[3]{n}$	1	1	1
$m = \sqrt[3]{n}$	1	3	3
$\sqrt[3]{n} < m < \sqrt{n}$	1	$\frac{2m^3}{n}$	$\frac{2m^3}{n}$
$m = \sqrt{n}$	3	$\frac{2\sqrt{n}}{3}$	$\frac{2\sqrt{n}}{9}$
$\sqrt{n} < m < n$	$\frac{2m^2}{n}$	m	$\frac{n}{2m}$
$m = n$	$\frac{14n}{27}$	n	$\frac{27}{14}$
$n < m$	$\frac{2m}{3}$	m	$\frac{3}{2}$

Section 3. Graphs of sh. paths

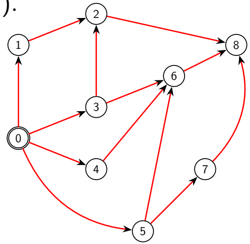
Graphs of sh. paths (GSP)

Given a weighted directed graph $G = (V, E, W)$:

Definition: Graph of Shortest Paths from s
Let s be a fixed node. The graph of shortest paths (GSP) from s is a directed graph $G_s = (V_s, E_s)$ defined as follows:



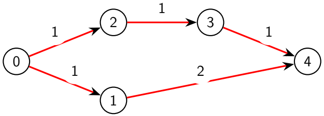
$$(i, j) \in E_s \iff (i, j) \in E \text{ and } d(s, j) = d(s, i) + W(i, j).$$



The set V_s corresponds to the nodes belonging to an edge from E_s which are the nodes accessible from s.
Remark: GSP is a directed acyclic graph that contains exactly one source.

Graphs of sh. paths (GSP) II

Prop: Any DAG containing one source is a GSP of some weighted directed graph.
Proof: add weights to DAG. Let s denote the source of the DAG and $m(s, w)$ the length of the longest path from s to w. The weight of the edge $u \to v$ denoted by $W(u, v)$ is set to $m(s, v) - d(s, u)$.

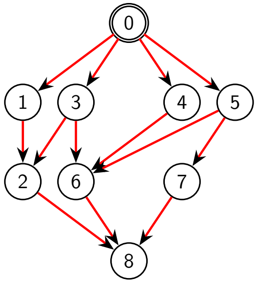


Graphs of sh. paths (GSP) III

Suppose G is unweighted. The GSP from s is such that:

- Contains one source s
- GSP is layered that is an edge (u, v) of the GSP goes from a node at distance $k = d(s, u)$ to a node at distance $k + 1 = d(s, v)$.

Consequence: GSP is bipartite and weakly connected



Notation: Profile is # nodes in each layer. On example $(4, 3, 1)$

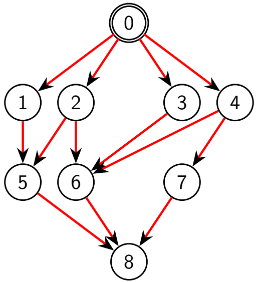
Section 4. Increasing GSPs

Increasing GSPs

Definition: A GSP with $n+1$ nodes is a layered DAG with one source. Nodes are labelled from 0 to n, the node labelled by 0 is the source. Then let ℓ_i be the number of nodes in layer i. Nodes in layer i are labelled from $\ell_1 + \cdots + \ell_{i-1} + 1$ to $\ell_1 + \cdots + \ell_i$.

Questions:

- Asymptotic Enum.
- Uniform Generation
- Limiting shape



Notation: $\ell(v)$ denotes the layer number of node v

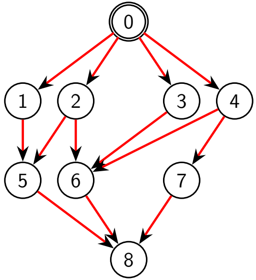
Increasing GSPs enum.

For a given profile $(\ell_1, \ldots, \ell_d)$, the number of GSP with this profile as layer sizes is:

$$f(\ell_1, \ldots, \ell_d) = \prod_{k=0}^{d-1} \left(2^{\ell_k} - 1\right)^{\ell_{k+1}}$$

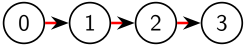
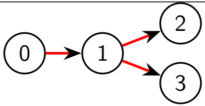
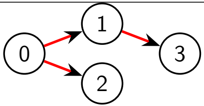
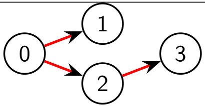
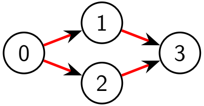
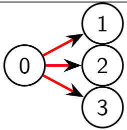
Therefore the total number of GSP with $n+1$ nodes is:

$$d_n = \sum_{\substack{(\ell_1,\ldots,\ell_d)\in\mathbb{N}^d \\ \ell_1+\cdots+\ell_d=n}} \prod_{k=0}^{d-1} \left(2^{\ell_k} - 1\right)^{\ell_{k+1}}.$$



$(d_n)_{n\geq 0} = (1, 1, 2, 6, 26, 158, 1330, 15486, 249922, 5604814, 175056146, \ldots)$

Increasing GSP of size 4

Profile	GSP	Profile	GSP
$(1, 1, 1)$		$(1, 2)$	
$(2, 1)$		$(2, 1)$	
$(2, 1)$		(3)	

Study by profile

- Summing all contributions of profiles of width 2 (those having 2 layers). Profiles: $(1, n-1), (2, n-2), \ldots (n-1, 1)$ gives $\sum_{k=1}^{n-1} (2^k - 1)^{n-k} = \Theta\left(2^{\frac{n^2}{4}}\right)$
- Take $k = \frac{n}{2}$ for lower bound and remove -1 in the sum for upper bound
- We also sum profiles of width 3

Definition: Let $n = 2k$ and $r \in [\![0; k-1]\!]$. We call **dominant profile of r-kind** the profiles having the form

- $(i, k-r, k+r-i)$ for $i \in [\![1; k+r]\!]$[1].
- $(i, k+r, k-r-i)$ for $i \in [\![1; k-r]\!]$.

We denote $d_{2k}^{(r)}$ the number of GSPs having a dominant profile of r-kind.

[1] The notation $[\![a; b]\!]$ stands for the set of integers from a to b.

Profiles of width 3

Lemma: Let $r > 0$. When $k \to +\infty$ we have $d_{2k}^{(r)} \geq 2k \cdot 2^{k^2 - r^2}(1 + o(1))$
Prop: Let $n = 2k$, when $k \to +\infty$, the following inequality holds:

$$d_n \geq S \cdot n \cdot 2^{\frac{n^2}{4}} \cdot (1 + o(1)) \quad \text{with} \quad S = \frac{1}{2} + \sum_{r=1}^{+\infty} \frac{1}{2^{r^2}}.$$

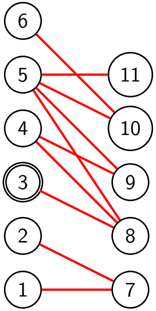
The same study can be made when n is odd. **Prop:** Let $n = 2k + 1$, when $k \to +\infty$ we have:

$$d_n \geq S'.n.2^{\frac{n^2}{4}}.(1 + o(1)) \quad \text{with} \quad S' = \sum_{r=0}^{+\infty} \frac{1}{2^{(r+\frac{1}{2})^2}}$$

Consequence: Profiles of width 2 are negligeable compared to those of width 3.

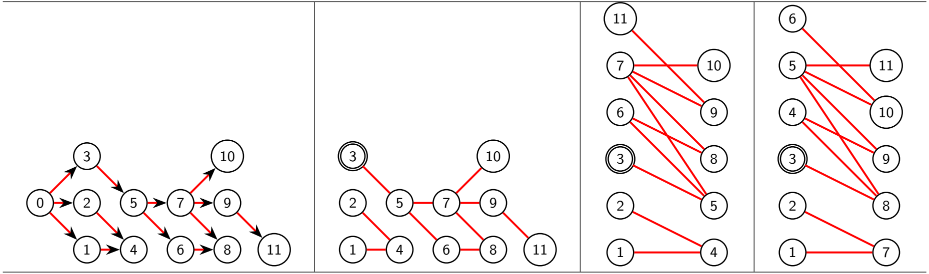
Marked bicolored graphs

For $n \in \mathbb{N}$, let's denote B_n the set of bicolored graphs having n nodes. We'll call $\mathcal{X}$ and $\mathcal{Y}$ the two partitions. In the partition $\mathcal{X}$ nodes are labelled from 1 to k and in the second nodes are labelled from $k+1$ to n. The partition $\mathcal{X}$ contains one marked node. Edges go only between nodes of different colors. We denote by $b_n = |B_n|$.



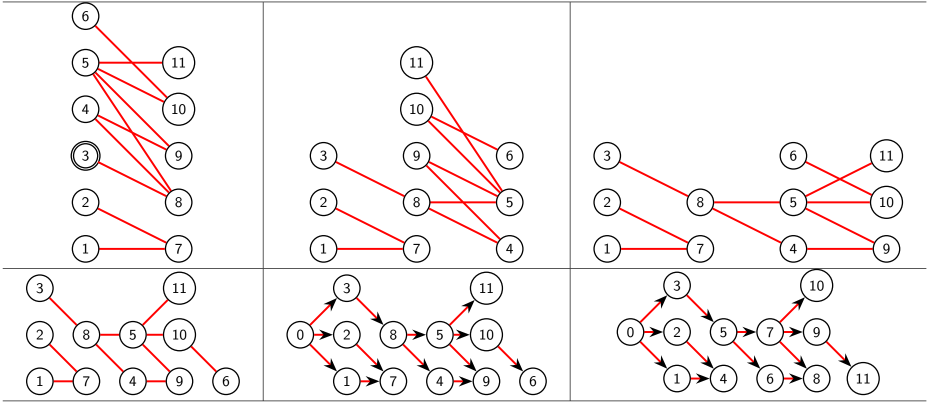
Folding

We give a construction $\mu : GSP \to MBG$ that transforms a GSP G with $n+1$ nodes into an MBG $\mu(G)$ with n nodes.



Unfolding

We define the partial reverse of $\mu^{-1} : MBG \rightharpoonup GSP$:



MBG Enumeration

Prop: b_n is an upper bound on d_n. (folding is injective)
We have $b_0 = 1$ (empty graph) and for $n \in \mathbb{N}^*$ the following holds:

$$b_n = \sum_{k=0}^{n} k.2^{k(n-k)}$$

Prop: Let $n \in \mathbb{N}^*$

$$b_n = \begin{cases} n.2^{\frac{n^2}{4}} \left(\frac{1}{2} + \sum_{r=1}^{\frac{n}{2}} \frac{1}{2^{r^2}} \right) & \text{when } n \text{ is even} \\ n.2^{\frac{n^2}{4}} \left(\sum_{r=0}^{\frac{n-1}{2}} \frac{1}{2^{(r+\frac{1}{2})^2}} \right) & \text{otherwise.} \end{cases}$$

Asymptotic Enumeration

From profiles of width 3:

$$d_n \geq \begin{cases} S \cdot n \cdot 2^{\frac{n^2}{4}} \cdot (1 + o(1)) \quad \text{with} \quad S = \frac{1}{2} + \sum_{r=1}^{+\infty} \frac{1}{2^{r^2}} & \text{when } n \text{ is even,} \\ S'.n.2^{\frac{n^2}{4}}.(1 + o(1)) \quad \text{with} \quad S' = \sum_{r=0}^{+\infty} \frac{1}{2^{(r+\frac{1}{2})^2}} & \text{otherwise.} \end{cases}$$

From marked bicolored graphs:

$$d_n \leq \begin{cases} n.2^{\frac{n^2}{4}} \left(\frac{1}{2} + \sum_{r=1}^{\frac{n}{2}} \frac{1}{2^{r^2}} \right) & \text{when } n \text{ is even} \\ n.2^{\frac{n^2}{4}} \left(\sum_{r=0}^{\frac{n-1}{2}} \frac{1}{2^{(r+\frac{1}{2})^2}} \right) & \text{otherwise.} \end{cases}$$

Theorem: The number of GSPs with $n+1$ nodes when n grows $+\infty$ is equivalent to:

For even n: $\quad d_n \sim n.2^{\frac{n^2}{4}} S \qquad$ with $S = \dfrac{1}{2} + \displaystyle\sum_{r=1}^{+\infty} \frac{1}{2^{r^2}}$

For odd n: $\quad d_n \sim n.2^{\frac{n^2}{4}} S' \qquad$ with $S' = \displaystyle\sum_{r=0}^{+\infty} \frac{1}{2^{(r+\frac{1}{2})^2}}$

Canonical Unfolding

Definition: Let $M = (\mathcal{X}, \mathcal{Y}, E)$ be an MBG with $\mathcal{X} = [x_1, \dots, x_k]$ and $\mathcal{Y} = [y_1, \dots, y_{n-k}]$ and G its incomplete unfolding then M is a canonical unfolding if the mappings $\hat{\mathcal{X}} = [\ell(x_1), \dots, \ell(x_k)]$ and $\hat{\mathcal{Y}} = [\ell(y_1), \dots, \ell(y_{n-k})]$ are increasing either by 0 or 2.

M_1 we have $\hat{\mathcal{X}} = [1, 3]$ and $\hat{\mathcal{Y}} = [2, 4]$ while for M_2, $\hat{\mathcal{X}} = [1, 3]$ and $\hat{\mathcal{Y}} = [4, 2]$. Therefore, M_2 is not a canonical unfolding of G.

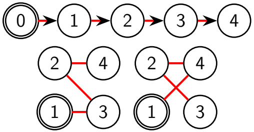


Figure: (up) a GSP G and (down) two MBG M_1 and M_2 succsessively having $\mathcal{X} = [1, 2]$ and $\mathcal{Y} = [3, 4]$. Applying μ^{-1} on M_1 and M_2 the right yields G and applying $\mu(G) = M_1$.

Uniform Sampling

Prop: Every GSP G has a unique canonical unfolding.

Uniform Sampling with rejection:

1. sample the partition size k according to the right distribution
2. sample a uniform bicolored graph M on with $|\mathcal{X}| = k$ and $|\mathcal{Y}| = n - k$
3. choose uniformly a node in $[1, \dots, k]$ and mark it
4. do the incomplete unfolding of M
5. If M is canonical return $\mu^{-1}(M)$ else start the whole process again

Consequence: From asymptotic enum. the rejection rate tends to 0 as n grows

Section 5. General GSPs

General GSPs

Definition: Layered DAG with one source with $n+1$ nodes labelled from 0 to n. The source is labelled 0.

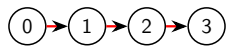
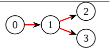
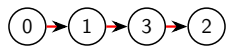
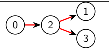
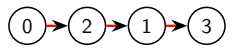
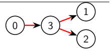
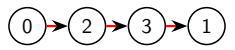
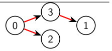
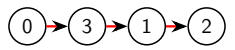
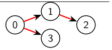
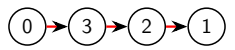
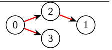
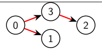
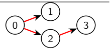
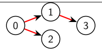
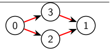
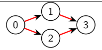
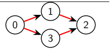
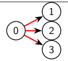
Also correponds to class of DAGs obtained from union DAGs that appear when taking all connected unweighted graphs of size n starting from node 0.

Let t_n be the number of general GSPs on $n+1$ nodes:

$$(t_n)_{n\geq 1} = 1, 1, 3, 19, 195, 3031, 67263, 2086099, 89224635, 5254054111, \ldots$$

The sequence t_n corresponds to $A001832$ in OEiS

General GSPs of size 4

Profile	GSP	Profile	GSP
(1, 1, 1)		(1, 2)	
(1, 1, 1)		(1, 2)	
(1, 1, 1)		(1, 2)	
(1, 1, 1)		(2, 1)	
(1, 1, 1)		(2, 1)	
(1, 1, 1)		(2, 1)	
(2, 1)		(2, 1)	
(2, 1)		(2, 1)	
(2, 1)		(2, 1)	
(3)			

General GSPs II

- Bijection with connected bipartite graphs [Wil05] and links with graded posets [Kla69]
- **Prop:** For even n,

$$c' 2^{\frac{n^2}{4}+\frac{3n}{2}} \frac{1}{\sqrt{n}} \leq t_n \leq c 2^{\frac{n^2}{4}+\frac{3n}{2}} \frac{1}{\sqrt{n}},$$

  with $c' = 2.020036\ldots$ and $c = 2.020041\ldots$
- Lower bound : profiles of width 3. Upper bound : using bicolored graphs
- **Conjecture:** $t_n \overset{n\to\infty}{=} c 2^{\frac{n^2}{4}+\frac{3n}{2}} \frac{1}{\sqrt{n}}$
- Uniform sampler : Sample uniformly and bicolored graph, if the resulting GSP is connected return it, else restart the whole process
- rejection rate if conjecture is true tends to 0 otherwise very small constant

SORBONNE UNIVERSITÉ LIP6

Bibliography I

[Boe17] Geoff Boeing. Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. Computers, environment and urban systems, 65:126–139, 2017.

[CM03] Aaron Clauset and Cristopher Moore. Traceroute sampling makes random graphs appear to have power law degree distributions. arXiv preprint cond-mat/0312674, 2003.

[CPBV14] Fabio Ciulla, Nicola Perra, Andrea Baronchelli, and Alessandro Vespignani. Damage detection via shortest-path network sampling. Physical review E, 89(5):052816, 2014.

[CT11] Christophe Crespelle and Fabien Tarissan. Evaluation of a new method for measuring the internet degree distribution: Simulation results. Computer Communications, 34(5):635–648, 2011.

[DAHB+06] Luca DallâAsta, Ignacio Alvarez-Hamelin, Alain Barrat, Alexei Vázquez, and Alessandro Vespignani. Exploring networks with traceroute-like probes: Theory and simulations. Theoretical Computer Science, 355(1):6–24, 2006.

[DOW24] Juan Dios Ortúzar and Luis G Willumsen. Modelling transport. John wiley & sons, 2024.

[FV07] Abraham D. Flaxman and Juan Vera. Bias reduction in traceroute sampling–towards a more accurate map of the internet. In International Workshop on Algorithms and Models for the Web-Graph, pages 1–15. Springer, 2007.

[Kla69] David A Klarner. The number of graded partially ordered sets. Journal of Combinatorial Theory, 6(1):12–19, 1969.

[Kun13] Jérôme Kunegis. Konect: the Koblenz network collection. In Proceedings of the 22nd international conference on world wide web, pages 1343–1350, 2013.

[LBCX03] Anukool Lakhina, John W. Byers, Mark Crovella, and Peng Xie. Sampling biases in ip topology measurements. In IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies, volume 1, pages 332–341. IEEE, 2003.

[LLFS07] Jérémie Leguay, Matthieu Latapy, Timur Friedman, and Kavé Salamatian. Describing and simulating internet routes. Computer Networks, 51(8):2067–2085, 2007.

Bibliography II

Graphs of sh. paths

[PFV10] Márton Pósfai, Attila Fekete, and Gábor Vattay. Shortest-path sampling of dense homogeneous networks. Europhysics Letters, 89(1):18007, 2010.

[Wil05] Herbert S Wilf. generatingfunctionology. CRC press, 2005.

[WVM10] Huijuan Wang and Piet Van Mieghem. Sampling networks by the union of m shortest path trees. Computer Networks, 54(6):1042–1053, 2010.

[ZZW+11] Guo-Qing Zhang, Shi Zhou, Di Wang, Gang Yan, and Guo-Qiang Zhang. Enhancing network transmission capacity by efficiently allocating node capability. Physica A: Statistical Mechanics and its Applications, 390(2):387–391, 2011.

SORBONNE UNIVERSITÉ LIP6