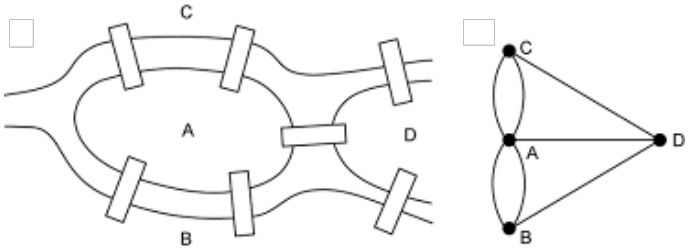
Targeted or lagged walk sampling for estimation of finite-order graph parameters

Li-Chun Zhang[1,2,3]

[1] ***Statistisk sentralbyraa, Norway***
[2] ***University of Southampton (L.Zhang@soton.ac.uk)***
[3] ***Universitetet i Oslo***

Graph, valued graph

Königsberg's bridges



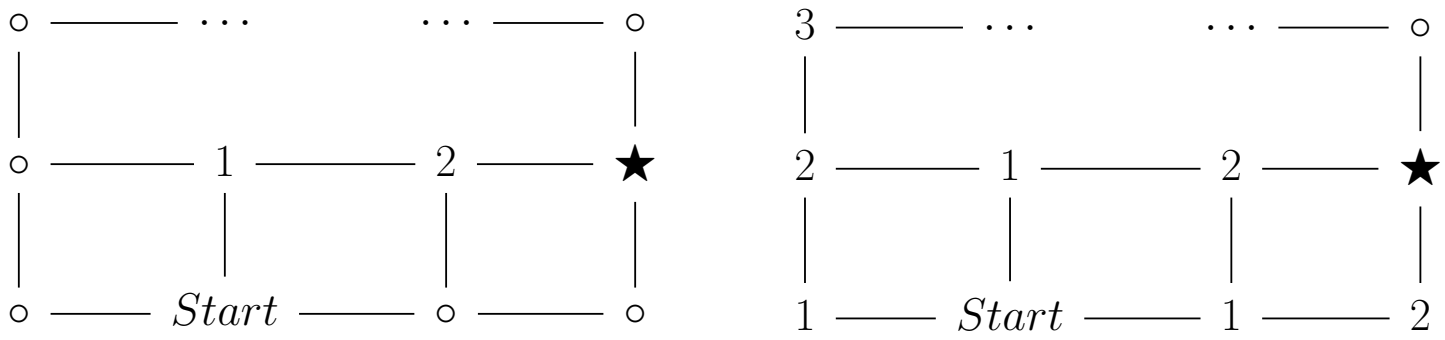
Euler's graph

Graph $G = (U, A) =$ (Nodes, Edges)
Nodes $U = \{i : i = 1, ..., N\}$ and N is the *order* of G
Edges $A = \{A_{ij} : i, j \in U\}$ and $|A|$ is the *size* of G
Simple graph if $A_{ij} = \{(ij)\}$ or $\emptyset$, *multigraph* otherwise

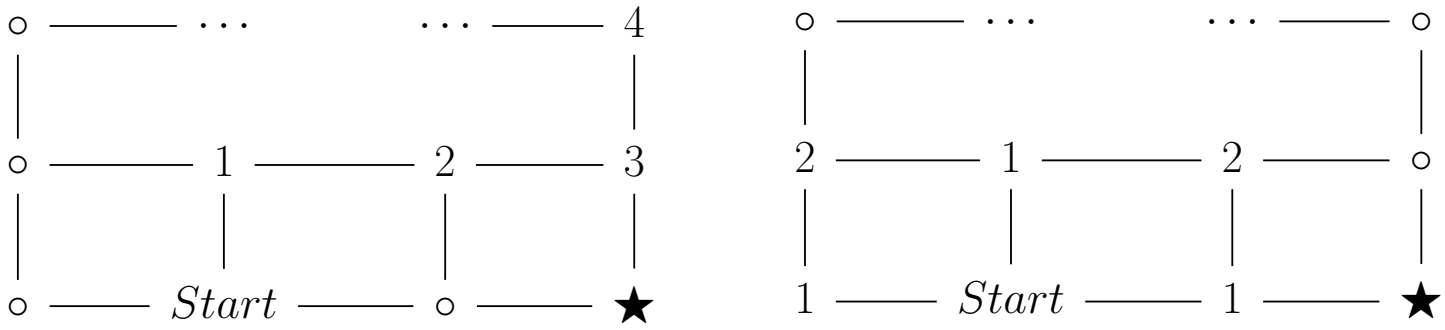
$A_{ij} =$ set of edges from i to j	nr. of edges $a_{ij} = \|A_{ij}\|$
out-edges $A_{i+} = \{A_{ij} : j \in U\}$	out-degree $a_{i+} = \|A_{i+}\|$
in-edges $A_{+i} = \{A_{ji} : j \in U\}$	in-degree $a_{+i} = \|A_{+i}\|$
undirected graph if $A_{ij} \equiv A_{ji}$	degree $d_i = a_{i+} = a_{+i}$

Attaching values to U or A yields a *valued graph*
Graph is the *structure* of a valued graph

Depth-first search vs. breadth-first search



Depth-first (left): if possible go up, right, down *or* left
Breadth-first (right): if possible, go up, right, down *and* left



↪ *Targeted random walk* ↪ *Snowball sampling*

Graph sampling methods: adopt strictly probabilistic rules (Zhang, 2021b)

Random walk (RW) in simple graphs $G = (U, A)$

Let $X_t = i$ be the node (or *state*) at step time t. Let

$$p_{ij} \coloneqq \Pr(X_{t+1} = j | X_t = i) = \frac{a_{ij}}{a_{i+}}$$

Select one out-edge randomly $(ij) \in A_{i+}$ yields $X_{t+1} = j$
Markov chain $\{X_t : t \geq 0\}$, transition probability matrix

$$P = [p_{ij}]$$

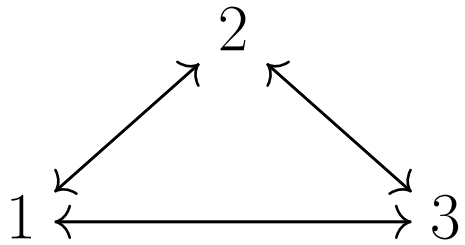
Stationary probability row-vector π at equilibrium:

$$\pi = \pi P \quad \text{and} \quad \pi_i \coloneqq \Pr(X_t = i),\ \forall i \in U$$

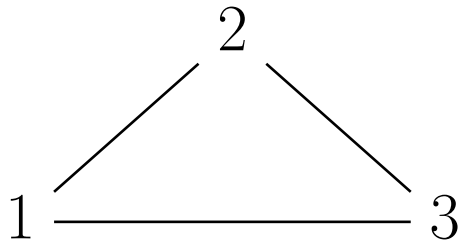
in particular, for connected undirected graphs,

$$\pi_i = \frac{d_i}{2|A|} = \frac{d_i}{\sum_{j \in U} d_j}$$

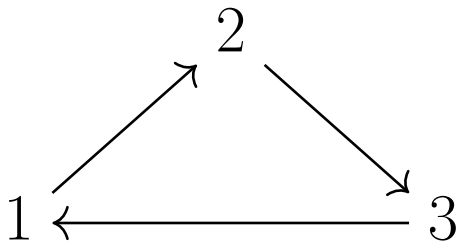
For directed graphs...



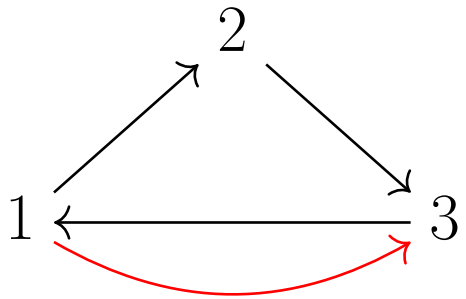
$\Leftrightarrow$



$$\pi_1 = \pi_2 = \pi_3 = \frac{1}{3}$$



$$\pi_1 = \pi_2 = \pi_3 = \frac{1}{3}$$



$$\begin{cases} \pi_1 = \pi_3 = \frac{2}{5} \\ \pi_2 = \frac{1}{2}\ \pi_1 = \frac{1}{5} \end{cases}$$

NB. $(a_{1+}, a_{+1}) = (2, 1)$ but $(a_{3+}, a_{+3}) = (1, 2)$

Random jumps

1 2 —— 3 need jumps to pass through 1

A proposal with random jumps

$$p_{ij} = \begin{cases} r\frac{a_{ij}}{a_{i+}} + (1-r)u_j & \text{if } a_{i+} > 0 \\ u_j & \text{if } a_{i+} = 0 \end{cases}$$

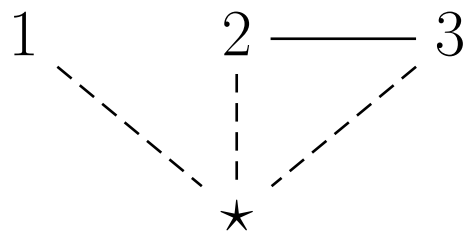
where r is called "damping factor" (Brin and Page, 1998), and $(u_1, ..., u_N)$ the preference vector, $\sum_{j \in U} u_j = 1$

If $u_j = a_{+j} / \sum_{i \in U} a_{+i}$ (e.g. Masuda et al. 2017), then 'PageRank'

$$\pi_i \approx \frac{a_{+i}}{\sum_{j \in U} a_{+j}} + \sum_{l=1}^{\infty} \frac{r^l}{\sum_{j \in U} a_{+j}} \sum_{j=1}^{N} (a_{+j} - a_{j+}) \left(\frac{a_{ji}}{a_{j+}}\right)^l$$

where j gives a positive contribution to i if j has a larger in-degree than out-degree, $a_{+j} > a_{j+}$, or negatively in the opposite case.

Targeted random walk (TRW)



Random jump as two successive moves via $\star$ (imaginary)

$$p_{ij} = \begin{cases} \frac{1}{d_i+r}\left(1+\frac{r}{N}\right) & \text{if } a_{ij}=1 \\ \frac{r}{d_i+r}\left(\frac{1}{N}\right) & \text{if } a_{ij}=0 \text{ including } i=j \end{cases}$$

NB. imaginary edge $(i\star)$ adds r to degree: $d_i \mapsto d_i + r$

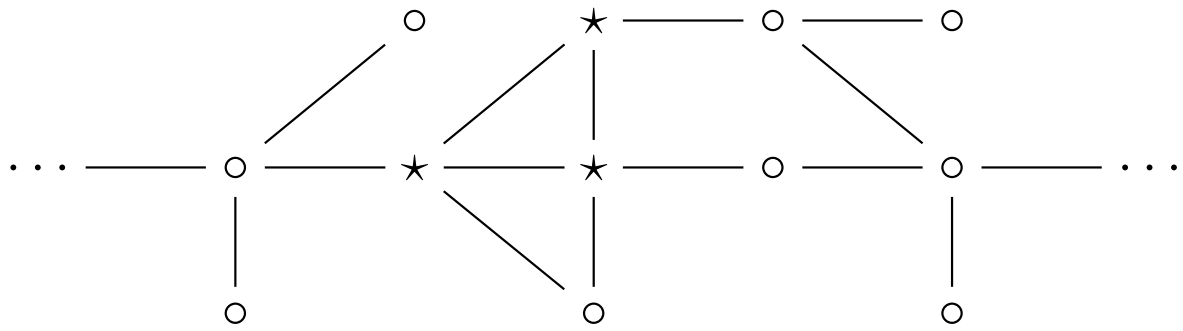
NB. probability of jump $\frac{r}{d_i+r}$ varies over nodes

For undirected graphs (Avrachenkov et al., 2010):

$$\pi_i \propto d_i + r$$

e.g. degree+1 walk if $r = 1$, almost pure RW if $r \approx 0$, almost (equal-π) uniform walk if $r \gg d_i$

Generalised ratio estimator (e.g. Thompson, 2006a)



Parameter: proportion $\star$ ($y = 1$) vs. $\circ$ ($y = 0$) over U

$$\mu = \frac{1}{N} \sum_{i \in U} y_i \quad \text{and} \quad N = |U|$$

Given known constants c_i, where $c_i \propto \pi_i$ by TRW,

$$\hat{\mu} = \Big(\frac{1}{n} \sum_{i \in \mathfrak{s}_n} \frac{y_i}{c_i} \Big) / \Big(\frac{1}{n} \sum_{i \in \mathfrak{s}_n} \frac{1}{c_i} \Big) = \sum_{i \in \mathfrak{s}_n} \frac{y_i}{c_i} / \sum_{i \in \mathfrak{s}_n} \frac{1}{c_i}$$

using an extraction of n states $\mathfrak{s}_n = \{X_{t_1}, X_{t_2}, ..., X_{t_n}\}$

NB. stationary draw-by-draw but not with-replacement sampling

Graph sampling theory generally

Representing a collection of units (U) by a graph (G) allows one to incorporate the connections (A) among the units in addition to the units themselves.

♢ The links may provide effectively access to those units that are the primary interest (sub-population of U)
e.g. indirect, network, adaptive cluster, line intercept (Zhang, 2021b)

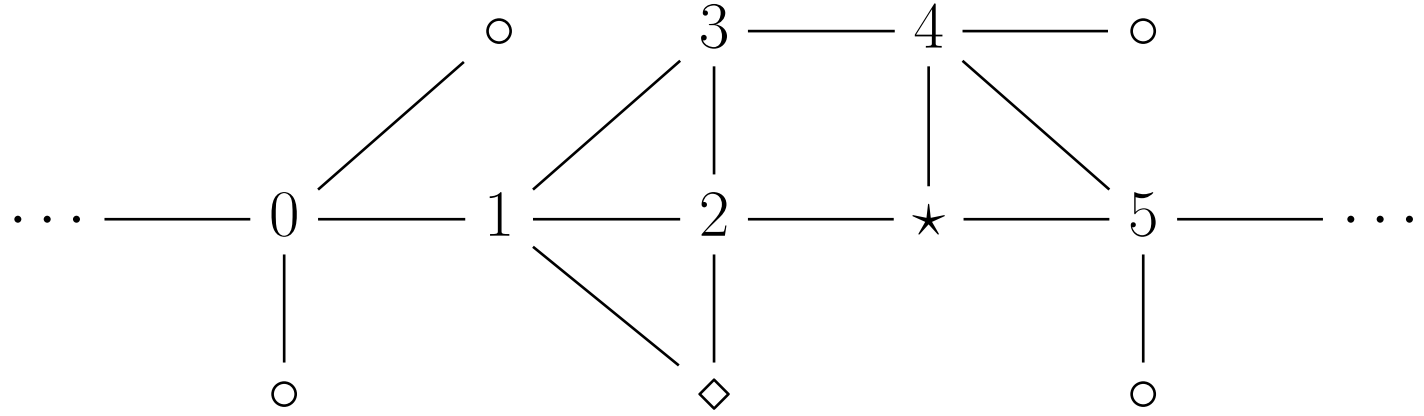
♡ The structure of the connections may be the interest of study — given as a graph problem to start with (e.g. Frank, 1971)

Graph sampling (Zhang, 2021b) analogous in concept to *finite population sampling* (Neyman, 1934)

- graph total/parameter – population total/parameter
- sample graph (subgraph) – sample (subpopulation)
- sampling strategy (sampling method, estimator)

♡ What do we observe other than $\{X_t : 0 \le t \le T\}$?

Under TRW, observe *all* edges incident to X_t at step t.



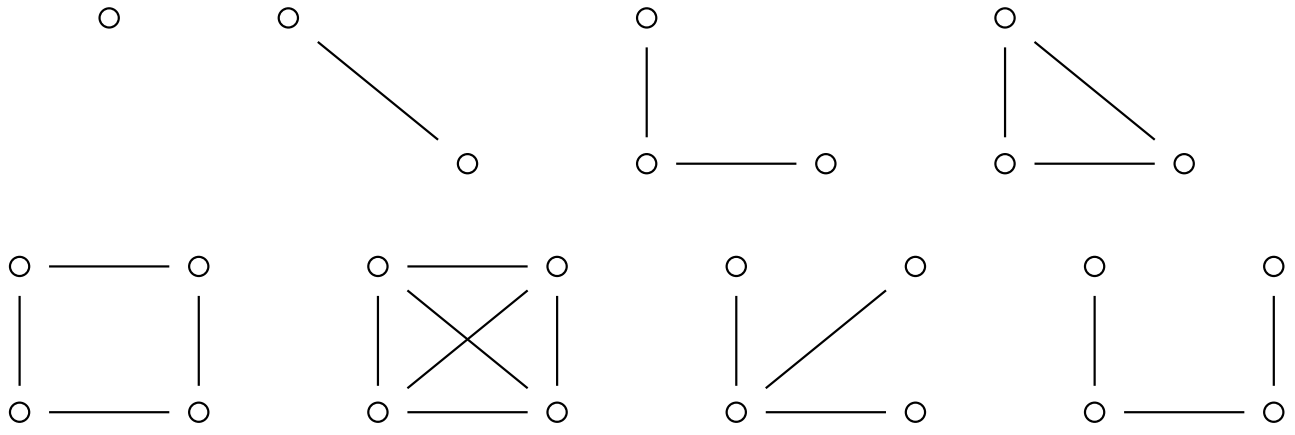
For T-step TRW sampling (TTRWS), let the *observation procedure (OP)* be applied to the *seed sample*

$$s = \{X_t : 0 \le t \le T\}$$

- observe edge of $\{0, 1\}$ at $X_t = 0$ and at $X_{t+1} = 1$
- triangle of $\{1, 2, 3\}$, $\{1, 2, \diamond\}$ given $(X_{t+1}, X_{t+2}) = (1, 2)$
  triangle of $\{1, 2, 3\}$ again given $(X_{t+2}, X_{t+3}) = (2, 3)$
- 4-cycle $\{2, 3, 4, \star\}$ given $(X_{t+2}, X_{t+3}, X_{t+4}) = (2, 3, 4)$

♡ Low-order motifs: Dyad, triad, tetrad

Frank (1971, 1977, 1978, 1979, 1980, 1981, 2011)



Node ($\mathcal{K}_1$), 2-clique ($\mathcal{K}_2$), 2-star ($\mathcal{S}_2$), 3-clique (triangle, $\mathcal{K}_3$), 4-cycle ($\mathcal{C}_4$), 4-clique ($\mathcal{K}_4$), 3-star ($\mathcal{S}_3$) and 3-path ($\mathcal{P}_3$)

$G(M) =$ subgraph *induced* by M, for $M \subset U$,
with nodes M and edges $\{A_{ij} : i, j \in M\}$
The specific characteristics of $G(M)$ is called
motif (Zhang, 2021b), the *order* of which is $|M|$.

Let $y(M)$ be a function of valued subgraph $G(M)$
Let Ω contain all the relevant node sets M
Graph total of $y(M)$ over Ω :

$$\theta = \sum_{M \in \Omega} y(M)$$

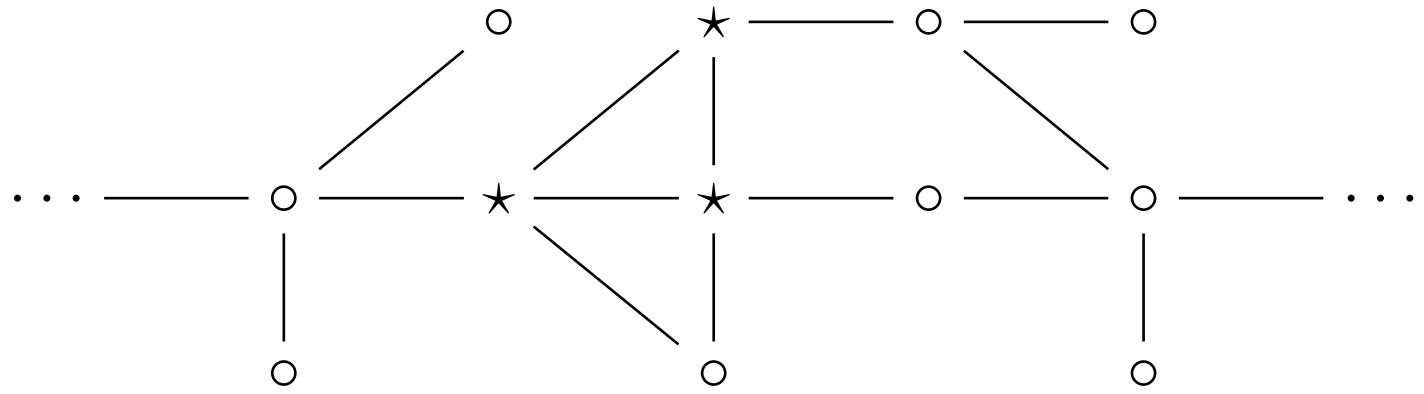
Graph parameter μ: any function of $\{y(M) : M \in \Omega\}$

Often convenient to let each $\kappa \in \Omega$ be an *occurrence* of a given motif (e.g. 2-star) in G, and write a graph total as

$$\theta = \sum_{\kappa \in \Omega} y_\kappa$$

and graph parameter μ is any function of $\{y_\kappa : \kappa \in \Omega\}$

An example of graph parameter



1st-order parameter of nodes $\star$ ($y = 1$) or $\circ$ ($y = 0$)

$$\mu_1 = \frac{1}{N} \sum_{i \in U} y_i \qquad \text{and} \qquad N = |U|$$

Now, e.g. triangle for graph transitivity or diffusion:

$$\mu_2 = \theta / \theta'$$

$\theta =$ total of triangles only consisting of $\star$

$\theta' =$ total of other triangles, not all $\star$-nodes

I. Sample graph G_s by TTRWS

Given seed sample $s = \{X_0, X_1, ..., X_T\}$ of TTRWS, the set of observed edges are given by

$$A_s = A \cap s_{\text{ref}} \quad \text{and} \quad s_{\text{ref}} = s \times U \cup U \times s$$

Sample graph is $G_s = (U_s, A_s)$ where $U_s = s \cup \text{Inc}(A_s)$

General definition (Zhang, 2021b, 2021a; Zhang & Patone, 2017): A method of sampling from $G = (U, A)$ has two parts:

- select an initial sample of nodes $s_0 \subset U$;
- given s_0, apply a specified OP making use of the edges in A.

The selected subgraph is a *sample graph* $G_s = (U_s, A_s)$, where

$$A_s = A \cap s_{\text{ref}} \qquad \text{and} \qquad U_s = s \cup \text{Inc}(A_s)$$

and s is the *seed sample* to which the OP has been applied, and s_{ref} specifies the observed part of the adjacency matrix $[a_{ij}]$.

II. Basis of inference: π_M for $M = \{X_{t_1}, ..., X_{t_q}\}$

Stationary probability at equilibrium $\pi_i = \Pr(X_t = i)$
Stationary sampling probability π_M may be unknown, e.g.

$$\pi_{X_t X_{t+2} X_{t+4}} = \pi_{135} = \pi_1 \Big(\sum_{i \in U} p_{1i} p_{i3} \Big) \Big(\sum_{i \in U} p_{3i} p_{i5} \Big)$$

Stationary successive sampling probability (S3P), e.g.

$$\pi_{X_t X_{t+1} X_{t+2} X_{t+3}} = \pi_{1234} = \pi_1 p_{12} p_{23} p_{34}$$

is known except for the proportionality constant in π_i
Generating states of T-step walk with seed sample s :

$$\mathcal{C}_s = \{M : M \subseteq s\}$$

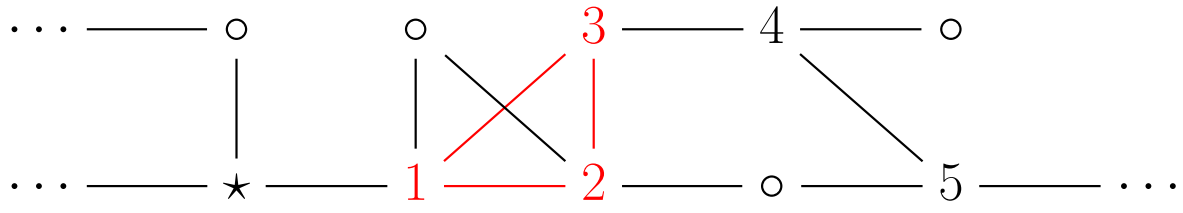
E.g. $(X_t, X_{t+1}, X_{t+2}) = (1, 2, 3)$ with S3P π_{123} actually... can also calculate S3P $\pi_{132}, \pi_{213}, \pi_{231}$ etc. hypothetically

III. Eligible sample motif κ if $\beta_\kappa \subset \mathcal{C}_s$

Actual sampling sequence of states (AS3) of motif κ:

$$s_\kappa = (X_t, ..., X_{t+q})$$

Equivalent sampling sequence of states (ES3) of s_κ, $\beta_\kappa = \{\tilde{s}_\kappa : \tilde{s}_\kappa \sim s_\kappa\}$, contains any possible sequence of states with $|\tilde{s}_\kappa| = |s_\kappa|$, such that the motif κ would be observed given $(X_t, X_{t+1}, ..., X_{t+q}) = \tilde{s}_\kappa$ but not based on any subsequence of $\tilde{s}_\kappa$. In particular, $s_\kappa \sim s_\kappa$.



Triangle κ of $M = \{1, 2, 3\}$ observed from
AS3 $(X_t, X_{t+1}) = (1, 2)$ as well as $(X_{t+1}, X_{t+2}) = (2, 3)$
ES3: $\beta_\kappa = \{(1, 2), (2, 1), (1, 3), (3, 1), (2, 3), (3, 2)\}$

III. Generalised ratio estimator (Zhang, 2021b)

Let AS3 $s_\kappa = (X_t, ..., X_{t+q})$ for eligible sample motif κ (in Ω)
Let M be a possible sequence of states $(X_t, ..., X_{t+q})$
Let $\delta_M = 1$ if M is realised and 0 otherwise, $\Pr(\delta_M = 1) = \pi_M$
Let $I_\kappa(M) = 1$ if $M \in \beta_\kappa$ and 0 otherwise (observation indicator)
Let $\{w_{M\kappa} : M \in \beta_\kappa\}$ be the *incidence weights*, $\sum_{M \in \beta_\kappa} w_{M\kappa} = 1$

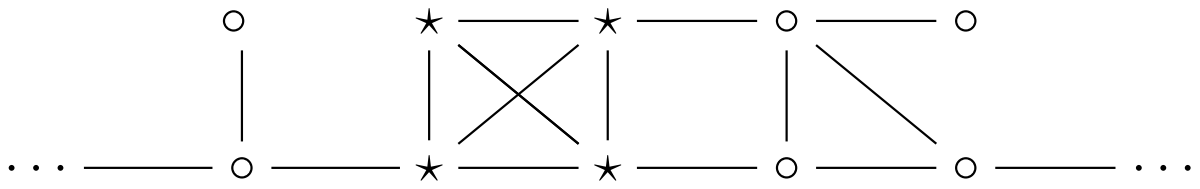
$$\hat{\theta}(X_t, ..., X_{t+q}) = \sum_{\kappa \in \Omega} \Big(\sum_{M} \frac{\delta_M}{\pi_M} I_\kappa(M) w_{M\kappa} \Big) y_\kappa$$

Given the AS3 s_κ is of the order $|s_\kappa| = q + 1$ and $|s| = n$, let
$\mathbb{I}_t = 1$ if $\sum_{\kappa \in \Omega} I_\kappa\big(\{X_t, ..., X_{t+q}\}\big) > 0$, and 0 otherwise

$$\hat{\theta} = \Big(\sum_{t=1}^{n-q} \mathbb{I}_t \hat{\theta}_t \Big) / \Big(\sum_{t=1}^{n-q} \mathbb{I}_t \Big)$$

Given $\theta = \sum_{\kappa \in \Omega} y_\kappa$ and $\theta' = \sum_{\kappa \in \Omega'} y'_\kappa$, can e.g. use $\boxed{\hat{\mu} = \hat{\theta}/\hat{\theta}'}$
if $\hat{\mu}$ invariant towards the unknown proportionality constant in π

An illustration (ψ = proportion of nodes visited)



Estimation of $\mu_1 = 0.2$ by TTRWS, 1000 simulations

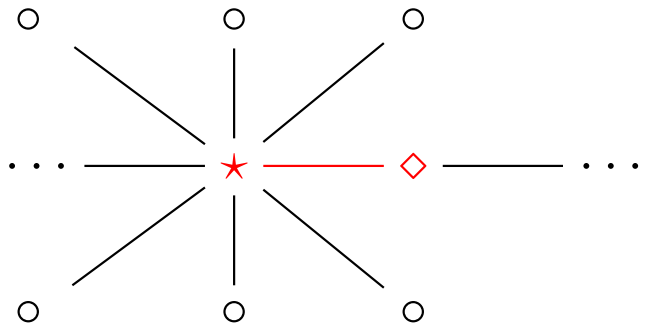
	$r = 1$			$r = 0.1$		
T	Mean($\hat{\mu}$)	SD($\hat{\mu}$)	ψ	Mean($\hat{\mu}$)	SD($\hat{\mu}$)	ψ
50	0.200	0.081	0.346	0.204	0.091	0.321
100	0.199	0.059	0.538	0.205	0.068	0.501
500	0.200	0.027	0.938	0.201	0.031	0.893
1000	0.199	0.019	0.987	0.201	0.022	0.959

Estimation of $\mu_2 = 4.667$ by TTRWS, 1000 simulations

	$r = 0.1$			$r = 6$		
T	Mean($\hat{\mu}$)	SD($\hat{\mu}$)	ψ	Mean($\hat{\mu}$)	SD($\hat{\mu}$)	ψ
100	6.119	4.498	0.498	6.362	5.069	0.606
500	4.737	0.893	0.893	4.805	1.075	0.983
1000	4.669	0.593	0.958	4.704	0.702	0.999

Non-Markovian lagged walk $\mapsto$ Markovian tuples

Isolated nodes may be of low interest... Consider the following:



Given $(X_{t-1}, X_t) = (\star, \diamond)$, should the move from $\diamond$ be *completely* random as in TRW? If no, X_{t+1} would *no longer* only depend on X_t.

Let *q-lagged random walk* $\{X_0, X_1, ..., X_T\}$ be given as

$$\underbrace{X_0, \overbrace{X_1, ..., X_q}^{\boldsymbol{x}_{q+1}}, X_{q+1}}_{\boldsymbol{x}_q}, ..., \underbrace{X_{t-q}, \overbrace{X_{t-q+1}, ..., X_t}^{\boldsymbol{x}_{t+1}}, X_{t+1}}_{\boldsymbol{x}_t}, ..., \underbrace{X_{T-q-1}, \overbrace{X_{T-q}, ..., X_{T-1}}^{\boldsymbol{x}_T}, X_T}_{\boldsymbol{x}_{T-1}}$$

$$\Pr(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, ..., \boldsymbol{x}_q) = \Pr(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t) = \Pr(X_{t+1}|\boldsymbol{x}_t)$$

NB. "A physical process... may or may not be Markovian, depending on the variables used to describe it." (Van Kampen, 1998)

An example of asymmetric RW (ARW), $q = 1$

$$1 \;\text{——}\; 2 \overset{p}{\text{——}} 3 \overset{1-p}{\text{——}} 4 \;\text{——}\; 5$$

ARW if $p \neq 0.5$, i.e. there is a tendency either to persist in the same direction or to backtrack to the previous state
NB. without jumps; $1 \to 2$ and $5 \to 4$ regardless p
We have

$$\begin{cases} \pi_2 = \pi_3 = \pi_4 \\ \pi_1 = \frac{1}{2}\pi_2 = \frac{1}{2}\pi_4 = \pi_5 \end{cases} \quad \Leftrightarrow \quad \pi_i \propto d_i$$

A difference is how quickly *all* the nodes are visited, e.g.

$E\big(\#\text{steps}\big) = 5.2$ if no backtracking at $X_t = 2, 3, 4$
$E\big(\#\text{steps}\big) \approx 9.5$ if $p = 0.5$

where the initial X_0 is randomly selected among $\{1, ..., 5\}$

ATRW sampling, generally

Let $\Pr\big(\boldsymbol{x}_{t+1} = (h, j) \mid \boldsymbol{x}_t = (i, h)\big)$ from $\boldsymbol{x}_t$ to $\boldsymbol{x}_{t+1}$ via h be

$$\begin{cases} \frac{r}{d_h+r}\left(\frac{1}{N}\right) + \frac{a_{hj}}{d_h+r} & \text{if } d_h = 1 \\ \frac{r}{d_h+r}\left(\frac{1}{N}\right) + \mathbb{I}(j = i)\frac{wa_{hj}}{d_h+r} + \mathbb{I}(j \neq i)\frac{d_h - wa_{ih}}{d_h+r}\frac{a_{hj}}{d_h - a_{ih}} & \text{if } d_h > 1 \end{cases}$$

If $(ih) \in A$ or $a_{ih} = 1$, then

- probability of backtracking given by $\Pr(j = i) \propto w$
- probability mass ($\propto d_h - w$) shared among other incident edges

Due to random jumps, probability $r/(d_h + r)$, ATRW is irreducible, so that a unique stationary distribution exists over $\boldsymbol{x}$, where

$$\pi_{\boldsymbol{x}} \coloneqq \Pr(\boldsymbol{x}_t = \boldsymbol{x}) \quad \text{and} \quad \sum_{\boldsymbol{x}} \pi_{\boldsymbol{x}} = 1$$

A unique stationary distribution of X_t follows as

$$\pi_h \coloneqq \Pr(X_t = h) = \sum_{g \in U} \Pr\big(\boldsymbol{x}_t = (g, h)\big), \quad \forall h \in U$$

ATRW sampling, generally

Lemma For ATRW in undirected simple graphs,

$$\pi_h \propto d_h + r, \quad \forall h \in U$$

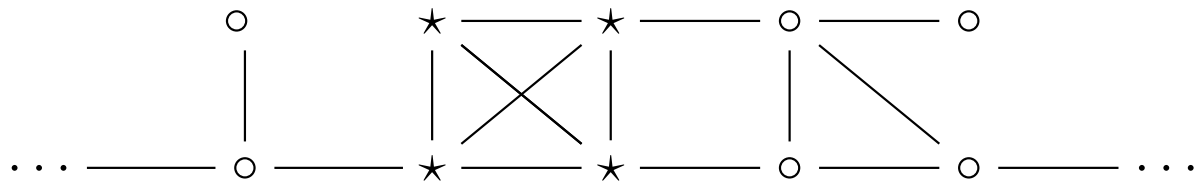
given any $0 \leq w \leq 1$, and for $\boldsymbol{x} = (i, h)$ we have

$$\pi_{\boldsymbol{x}} \propto \begin{cases} 1 + r/N & \text{if } a_{ih} = 1 \\ r/N & \text{if } a_{ih} = 0 \end{cases}$$

NB. a flow between $\boldsymbol{x}_t = (i, h)$ and $\boldsymbol{x}_{t+1} = (h, j)$ in either direction is a flow over (X_{t-1}, X_t, X_{t+1}) in the same direction.
NB. at equilibrium, the values $\{p_h : h \in U\}$ satisfying balanced flows through (X_{t-1}, X_t, X_{t+1}) would yield $\pi_h \propto p_h$ for any $h \in U$.
NB. can build a (non-straightforward) proof on π_h given as

$$\pi_h = \sum_{\substack{\boldsymbol{x}=(i,h) \\ i \in \nu_h}} \pi_{\boldsymbol{x}} + \sum_{g \notin \nu_h} \frac{\pi_g}{d_g + r} \left(\frac{r}{N}\right)$$

Illustration revisited



Estimation of $\mu_1 = 0.2$ by (A)TRWS, 1000 simulations

	$r = 0.1$			$r = 1$		
$T = 100$	Mean($\hat{\mu}$)	SD($\hat{\mu}$)	ψ	Mean($\hat{\mu}$)	SD($\hat{\mu}$)	ψ
$w = 1$	0.205	0.068	0.501	0.199	0.059	0.538
$w = 0.01$	0.204	0.056	0.559	0.204	0.053	0.577

Estimation of $\mu_2 = 4.667$ by (A)TRWS, 1000 simulations

	$r = 0.1$			$r = 6$		
$T = 100$	Mean($\hat{\mu}$)	SD($\hat{\mu}$)	ψ	Mean($\hat{\mu}$)	SD($\hat{\mu}$)	ψ
$w = 1$	6.119	4.498	0.498	6.362	5.069	0.606
$w = 0.01$	5.593	3.337	0.560	6.021	4.858	0.619
$T = 500$	Mean($\hat{\mu}$)	SD($\hat{\mu}$)	ψ	Mean($\hat{\mu}$)	SD($\hat{\mu}$)	ψ
$w = 1$	4.737	0.893	0.893	4.805	1.075	0.983
$w = 0.01$	4.724	0.785	0.920	4.767	1.005	0.984

Zhang (2021b) by CRC Focus, to appear Nov. 2021

"New and Emerging" general graph sampling theory

♡ graph total/parameter defined for finite-order subgraphs

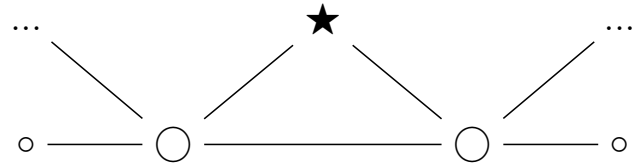
♡ breadth- or depth-first probability sampling algorithms in the form of TSBS and TRWS, respectively

♡ generally applicable graph sampling strategies

◇ Encompasses all finite-population sampling techniques

Some directions for future development

- other "sampling techniques", hybrids of depth-breadth-first, e.g. AWS (Thompson, 2006b) — scalable methods needed
- other graph parameters, e.g.



- other observation procedures, e.g. enabled by data structure

"Graph sampling is clearly the future of sampling."

— Zhang (2021a)

[1] Avrachenkov, K., Ribeiro, B. and Towsley, D. (2010). Improving Random Walk Estimation Accuracy with Uniform Restarts. *Research report*, RR-7394, INRIA. inria-00520350

[2] Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30:107-117.

[3] Frank, O. (1971). *Statistical inference in graphs*. Stockholm: Försvarets forskningsanstalt.

[4] Frank, O. (1977). Survey sampling in graphs. *Journal of Statistical Planning and Inference*, 1:235-264.

[5] Frank, O. (1978). Estimation of the number of connected components in a graph by using a sampled subgraph. *Scandinavian Journal of Statistics*, 5:177-188.

[6] Frank, O. (1980). Sampling and inference in a population graph. *International Statistical Review*, 48:33-41.

[7] Frank, O. (1981). A survey of statistical methods for graph analysis. *Sociological methodology*, 12:110-155.

[8] Frank, O. (2011). Survey sampling in networks. *The SAGE Handbook of Social Network Analysis*, pp. 389-403.

[9] Masuda, N., Porter, M.A. and Lambiotte, R. (2017) Random walks and diffusion on networks. *Physics Reports*, 716-717: 1-58. `http://dx.doi.org/10.1016/j.physrep.2017.07.007`

[10] Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97:558-606.

[11] Thompson, S.K. (2006a). Targeted random walk designs. *Survey Methodology*, 32, 11–24.

[12] Thompson, S.K. (2006b). Adaptive Web Sampling. *Biometrics*, 62, 1224–1234.

[13] Van Kampen, N.G. (1998). Remarks on Non-Markov Processes. *Brazilian Journal of Physics*, 28:90-96.

[14] Zhang, L.-C. (2021a). Graph sampling: An introduction. *The Survey Statistician*, 83:27-37.

[15] Zhang, L.-C. (2021b). *Graph sampling*. CRC Focus, to appear Nov. 2021.

[16] Zhang, L.-C. and Patone, M. (2017). Graph sampling. *Metron*, 75:277-299.