Sampling regular directed graphs in polynomial time

Catherine Greenhill

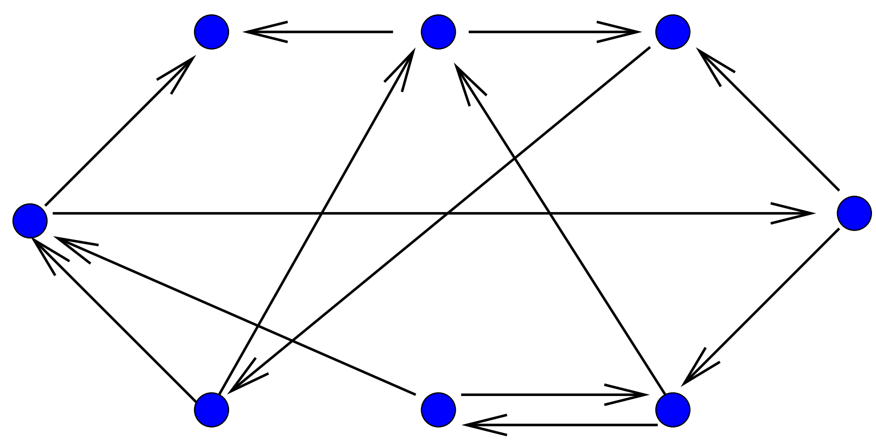
School of Mathematics and Statistics

University of New South Wales, Australia

(Currently on sabbatical at the
University of Durham, UK, until end June 2012)

A directed graph (or digraph) $G = (V, A)$ consists of a set V of vertices and a set A of arcs, where each arc is an ordered pair of distinct vertices (v, w).

Our digraphs are finite, so assume that $V = [n] = \{1, 2, \ldots n\}$.

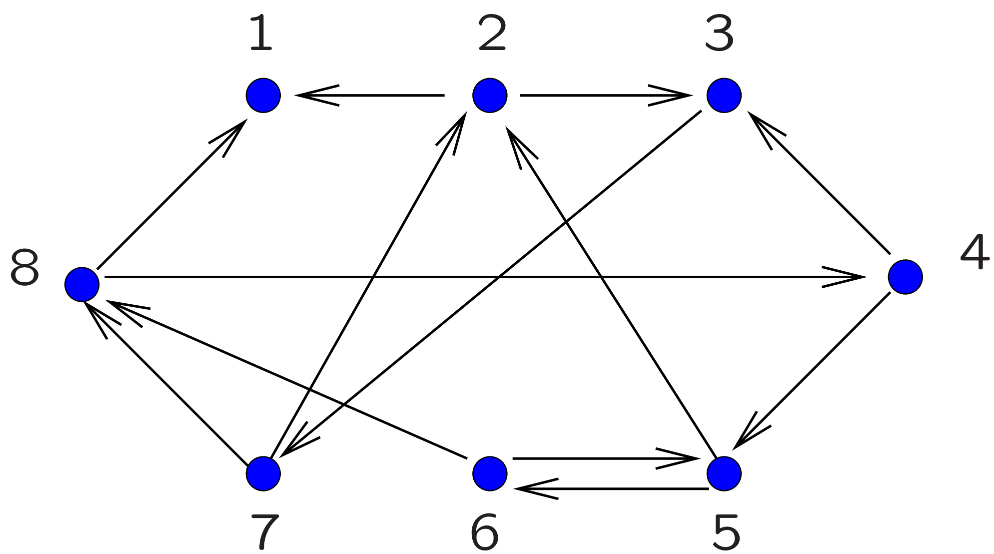


Let v be a vertex in a digraph G. The in-degree of v in G is the number of arcs $(w, v) \in A$ which terminate at v, while the out-degree of v is the number of arcs $(v, w) \in A$ which originate at v.

Given two vectors of nonnegative integers $\boldsymbol{d}^- = (d_1^-, \ldots, d_n^-)$ and $\boldsymbol{d}^+ = (d_1^+, \ldots, d_n^+)$ with the same sum, let $\mathcal{S}(n, \boldsymbol{d}^-, \boldsymbol{d}^+)$ be the set of all directed graphs with vertex set $[n]$ such that vertex i has in-degree d_i^- and out-degree d_i^+ for all $i \in [n]$.

Note: the entries of $\boldsymbol{d}^-, \boldsymbol{d}^+$ may depend on n.

Here $\boldsymbol{d}^- = (1, 2, 2, 1, 2, 1, 1, 2)$ and $\boldsymbol{d}^+ = (0, 2, 1, 2, 2, 1, 2, 2)$:



In many applications we would like an efficient algorithm for sampling uniformly from $\mathcal{S}(n, \boldsymbol{d}^-, \boldsymbol{d}^+)$.

Sampling digraphs with fixed degrees

Polynomial time means in time poly$(n, d_{\max})$ where $d_{\max} = \max\{d_1^-, \ldots, d_n^-, d_1^+, \ldots, d_n^+\}$.

• The configuration model (Bollobás, 1980) performs uniform sampling in expected polynomial time if $d_{\max} = O(\sqrt{\log n})$.

• An algorithm of McKay & Wormald (1990) can be adapted to perform uniform sampling in expected polynomial time if $d_{\max} = O(\log n)$.

I know of no other efficient uniform sampling algorithms for $\mathcal{S}(n, \boldsymbol{d}^-, \boldsymbol{d}^+)$. So, we will try approximately uniform sampling in (deterministic) polynomial time using a Markov chain.

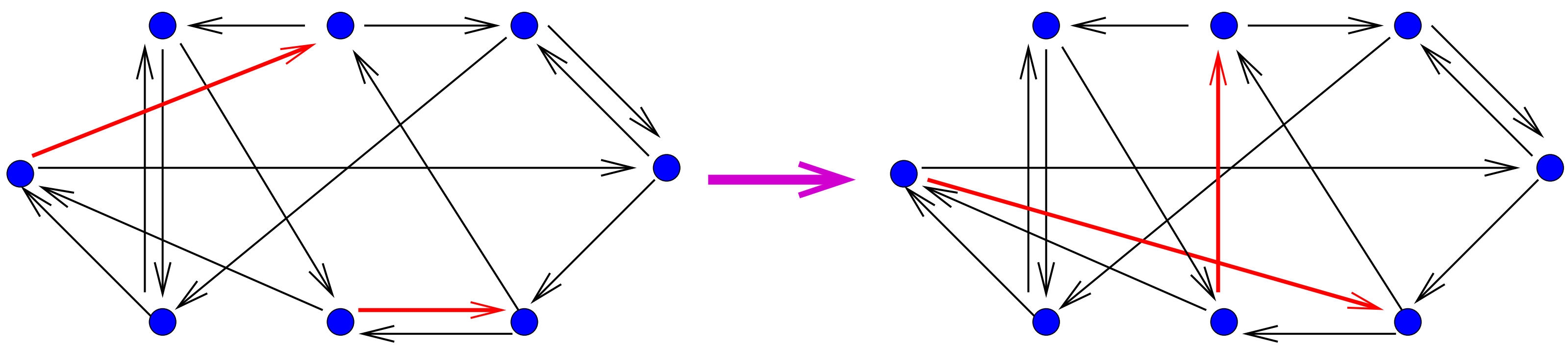
Related work

Kim, Del Genio, Bassler & Toroczkai (2012) gave a polynomial-time algorithm for sampling directed graphs with fixed in- and out-degrees, from a specific, computable, non-uniform distribution. (They can also do exhaustive generation.)

Then biased sampling can be used to calculate (unweighted) averages of various statistics.

I think we will hear more about this after coffee.

A very natural Markov chain on $\mathcal{S}(n, \boldsymbol{d}^-, \boldsymbol{d}^+)$ uses switches. We call this chain the switch chain.



From $G \in \mathcal{S}(n, \boldsymbol{d}^-, \boldsymbol{d}^+)$ do

choose an unordered pair of distinct arcs $\{(i,j),\ (k,\ell)\} \subseteq A(G)$ uniformly at random;

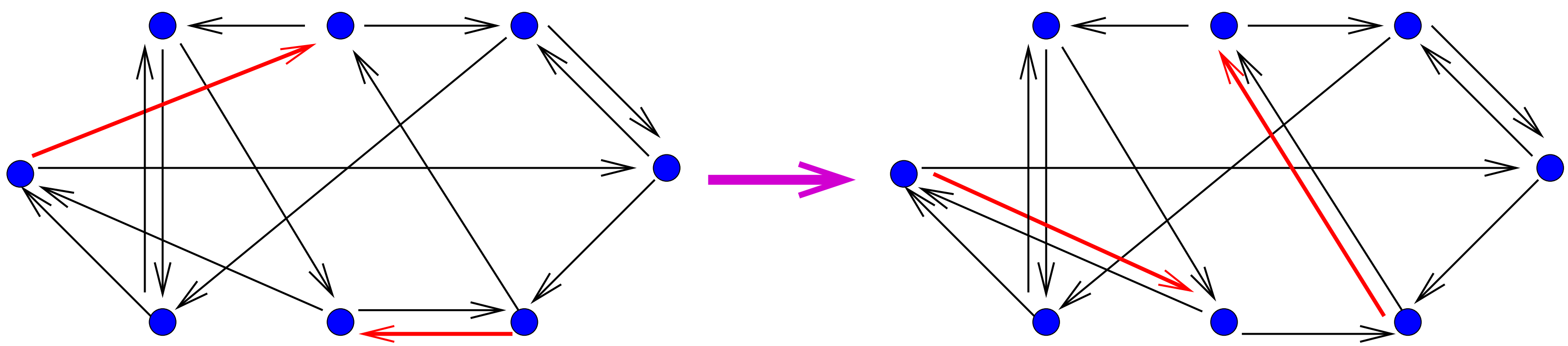
if $|\{i,j,k,\ell\}| = 4$ and $\{(i,\ell),(k,j)\} \cap A(G) = \emptyset$ then

replace these arcs with $\{(i,\ell),\ (k,j)\}$;

else

do nothing.

A very natural Markov chain on $\mathcal{S}(n, \boldsymbol{d}^-, \boldsymbol{d}^+)$ uses switches.
We call this chain the switch chain.



From $G \in \mathcal{S}(n, \boldsymbol{d}^-, \boldsymbol{d}^+)$ do

choose an unordered pair of distinct arcs

$\{(i,j),\ (k,\ell)\} \subseteq A(G)$ uniformly at random;

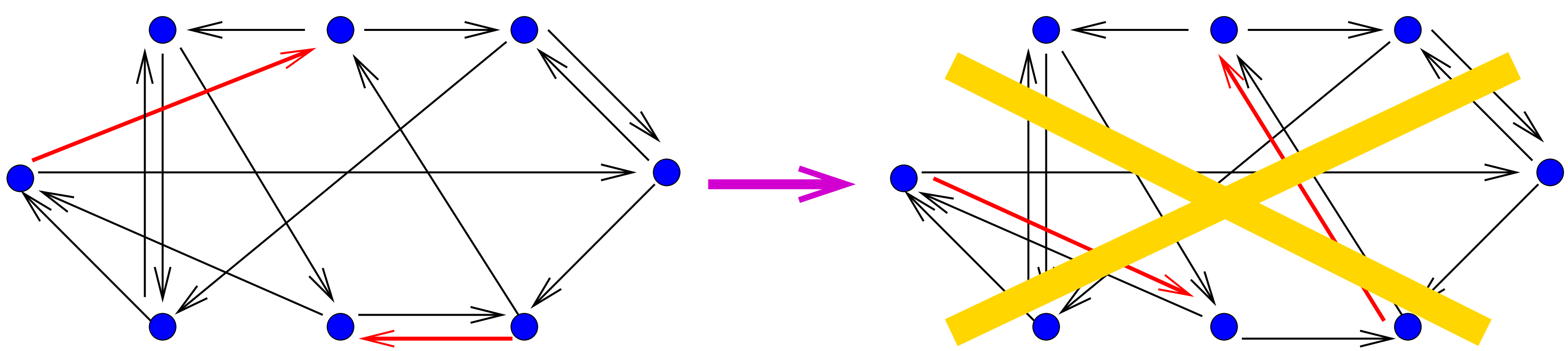
if $|\{i,j,k,\ell\}| = 4$ and $\{(i,\ell),(k,j)\} \cap A(G) = \emptyset$ then

replace these arcs with $\{(i,\ell),\ (k,j)\}$;

else

do nothing.

A very natural Markov chain on $\mathcal{S}(n, \boldsymbol{d}^-, \boldsymbol{d}^+)$ uses switches.
We call this chain the switch chain.



From $G \in \mathcal{S}(n, \boldsymbol{d}^-, \boldsymbol{d}^+)$ do

choose an unordered pair of distinct arcs

$\{(i,j),\ (k,\ell)\} \subseteq A(G)$ uniformly at random;

if $|\{i,j,k,\ell\}| = 4$ and $\{(i,\ell),(k,j)\} \cap A(G) = \emptyset$ then

replace these arcs with $\{(i,\ell),\ (k,j)\}$;

else

do nothing.

Ryser (1963) used switches to study 0-1 matrices.

Markov chains based on switches have been introduced by

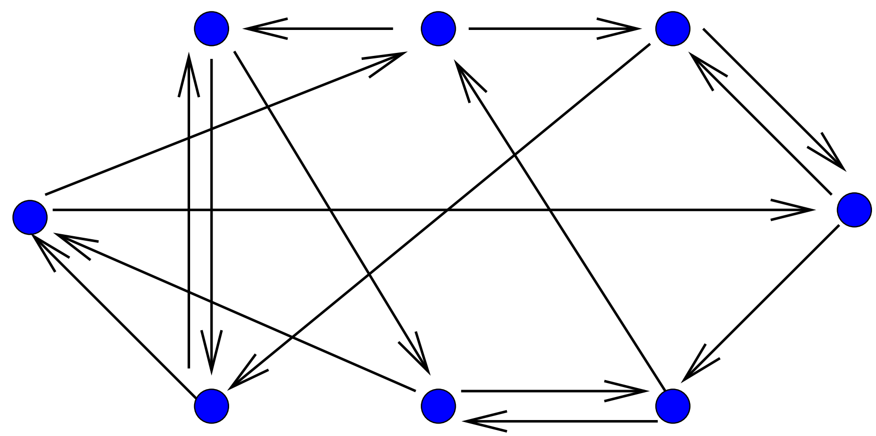
* Besag & Clifford (1989), for 0-1 matrices,

* Diaconis and Sturmfels (1995) and Holst (1995), for contingency tables,

* Rao, Jana & Bandyopadhyay (1996), for digraphs.

Restrict to regular digraphs

If every vertex $v \in V$ has in-degree d and out-degree d then we say that G is d-regular (or d-in, d-out).



Let $\mathcal{S}_{n,d}$ be the set of all d-regular digraphs on the vertex set $[n]$. Here $d = d(n)$ might depend on n, and satisfies $1 \leq d(n) \leq n - 1$ for all n.

Rao, Jana & Bandyopadhyay (1996) showed that the switch chain is not always irreducible on $\mathcal{S}(n, \boldsymbol{d}^-, \boldsymbol{d}^+)$, but that you obtain an irreducible Markov chain if you reverse a directed 3-cycle occasionally.

LaMar (2009) gave a characterisation of degree sequences $(\boldsymbol{d}^-, \boldsymbol{d}^+)$ for which the switch chain is irreducible. (See also Berger & Müller-Hannemann 2009.)

It follows from this characterisation that the switch chain is irreducible on $\mathcal{S}_{n,d}$.

The switch chain is aperiodic and its stationary distribution is uniform.

In 2011 I proved that the switch chain on $\mathcal{S}_{n,d}$ converges to within ε of the uniform distribution (in total variation distance) after at most

$$50d^{25}n^9(dn \log(dn) + \log(1/\epsilon))$$

steps. The analysis used a multicommodity flow argument, building on the undirected case (Cooper, Dyer & Greenhill, 2007).

Main steps:

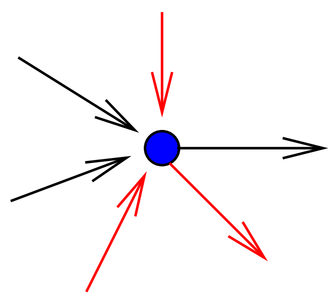
• For each $X \neq Y \in \mathcal{S}_{n,d}$, define a set of paths from X to Y, where each step is a transition of the switch chain.

• Analyse the congestion of the set of all paths: are any transitions heavily loaded? Then apply Sinclair (1992).

Defining the flow

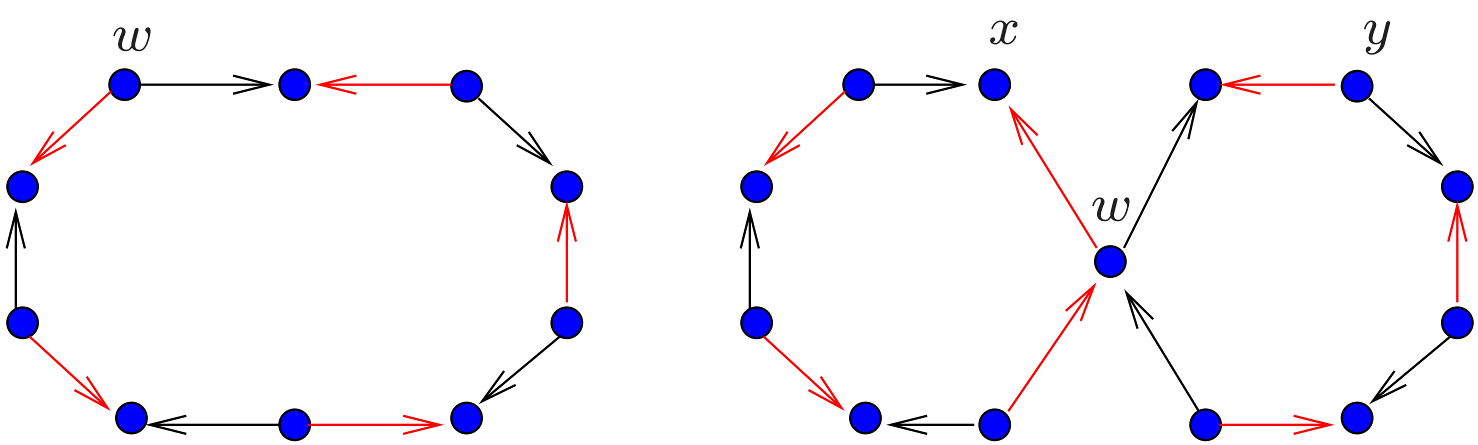
Given $X \neq Y \in \mathcal{S}_{n,d}$, consider the symmetric difference H of X and Y. Colour $X - Y$ black and $Y - X$ red.

For each vertex $v \in [n]$, pair up each in-arc at v with an in-arc of a different colour, and similarly for out-arcs. This gives a pairing of H.



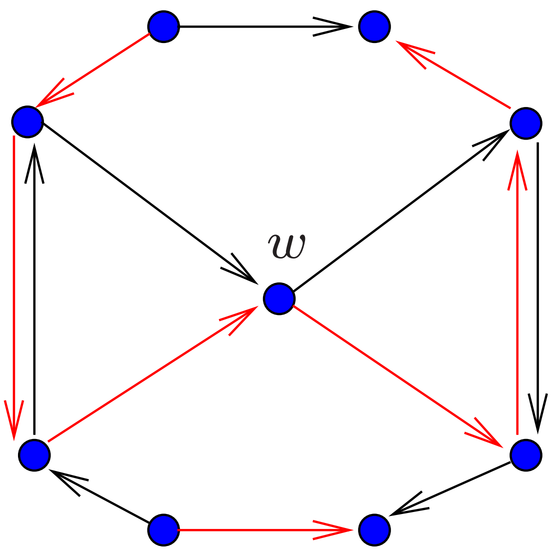
We define a path $\gamma_\psi(X, Y)$ from X to Y for each pairing ψ of H.

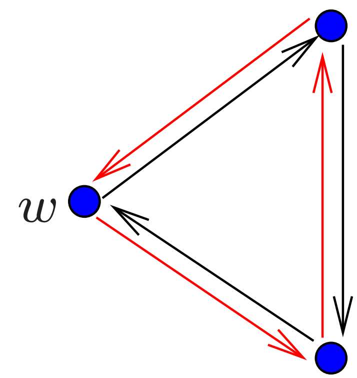
First we pull H apart into a sequence of 1-circuits and 2-circuits, following ψ. Here w is the start vertex which is traversed exactly once on a 1-circuit, exactly twice on a 2-circuit.



These can be processed as in CDG (2007) unless $x = y$.

We have to deal with some grisly 2-circuits that do not arise in the undirected case:

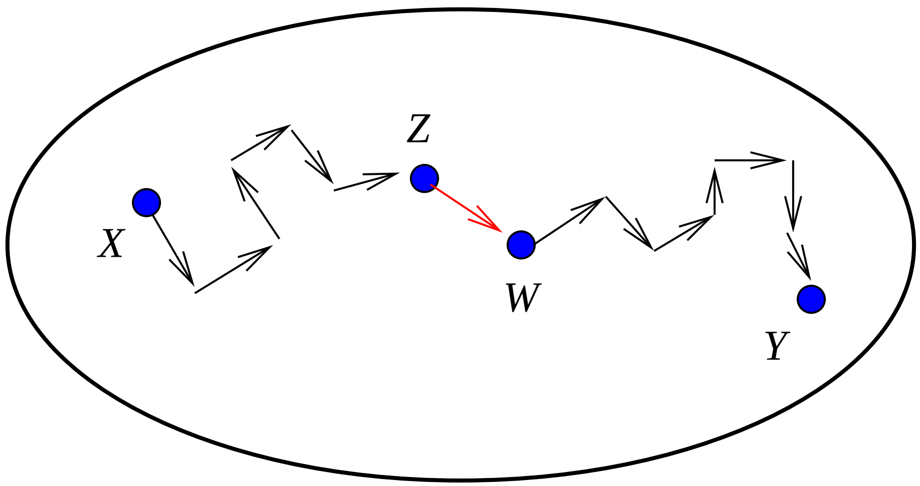




But these can be handled, by extending the argument from CDG (2007) and using results from LaMar (2009) for the triangle.

Analysing the flow:

Let (Z, W) be a transition which occurs on a path $\gamma_\psi(X, Y)$ from X to Y.



How much information do you need to uniquely reconstruct X and Y from (Z, W, ψ)?

Identify elements of $\mathcal{S}_{n,d}$ with their $n \times n$ adjacency matrices and let

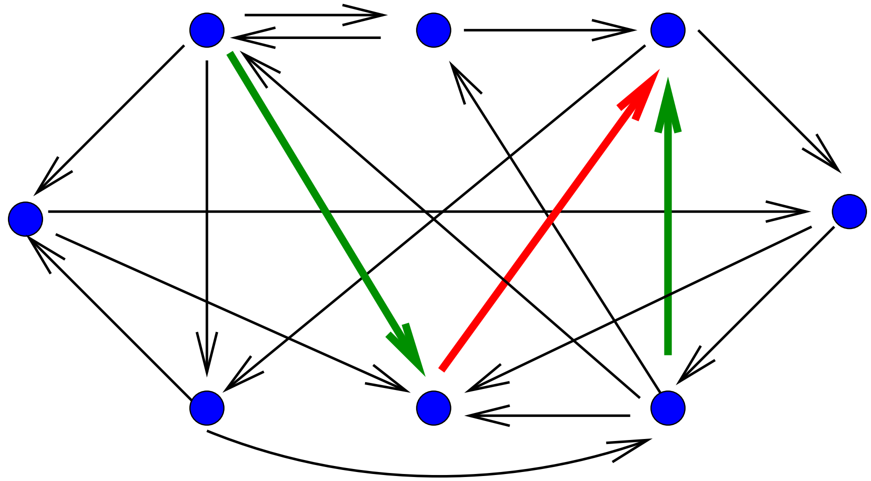
$$L = X + Y - Z.$$

The matrix L is called an encoding. Note, every row of L sums to d, and the same for the columns. Entries of L belong to $\{-1, 0, 1, 2\}$ and entries not equal to 0 or 1 are called defects.

A defect entry of -1 corresponds to an arc which is present in Z but absent in both X and Y.

A defect entry of 2 corresponds to an arc which is absent in Z but present in both X and Y.

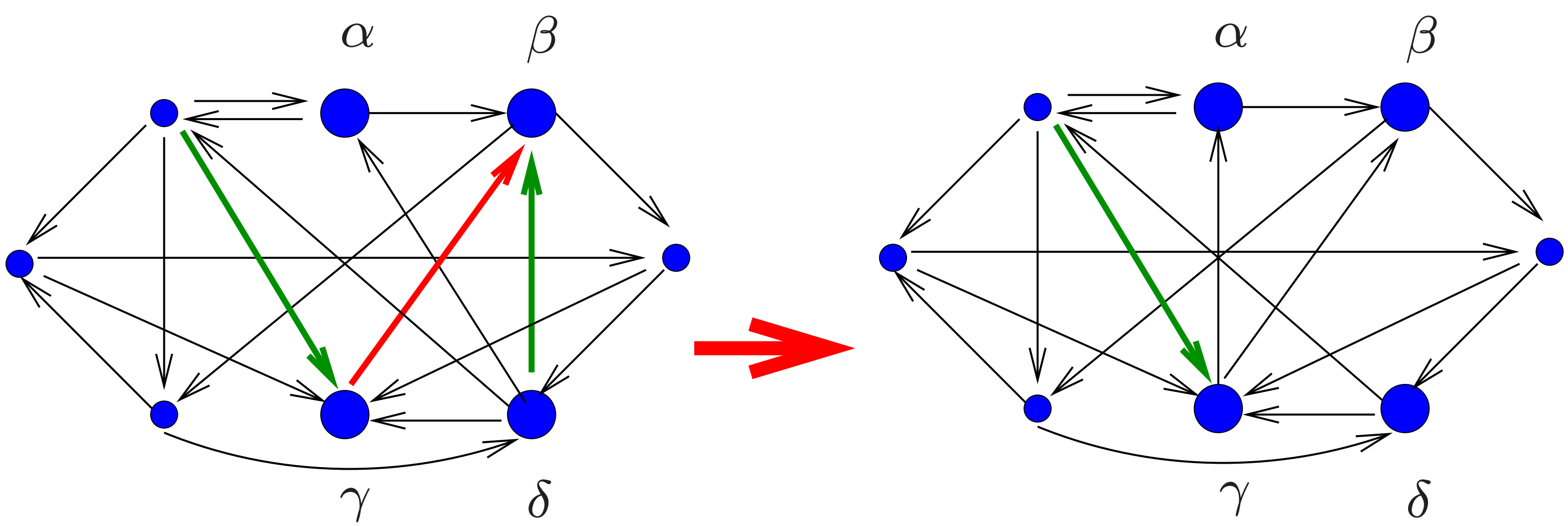
An encoding is shown below: red arcs are labelled 2 and green arcs are labelled -1.



Fact: Given (Z, W, ψ, L), there are at most four choices for (X, Y) such that $(Z, W) \in \gamma_\psi(X, Y)$.

Next we must show that there are at most $\text{poly}(n, d)\, |\mathcal{S}_{n,d}|$ encodings.

Critical Fact: at most three switches are needed to move from an arbitrary encoding to an element of $\mathcal{S}_{n,d}$.



This follows since there are at most 5 defects in any encoding, and the defects satisfy some other structural properties.

What about irregular degree sequences?

- First check that the switch chain is irreducible for the given in- and out-degrees using LaMar (2009);

- We can define the multicommodity flow exactly as in the regular case;

- Many steps of the analysis go through unchanged. But it is no longer clear that every encoding is within some small number of switches of a defect-free digraph.

This is a serious problem!

Questions/Future work:

- Can the regularity condition be relaxed at all? In the undirected case Erdős, Miklós and Soukup (arXiv, 2010) show that the undirected switch chain for bipartite graphs is efficient so long as the degrees on one side of the vertex bipartition are regular.

- Bayati, Kim & Saberi (2009) presented a sequential importance sampling algorithm for sampling undirected graphs with fixed degrees almost uniformly. Their algorithm is efficient if $d_{\max} = o(m^{1/4})$ (but with a small failure probability). Adapt this for directed graphs?