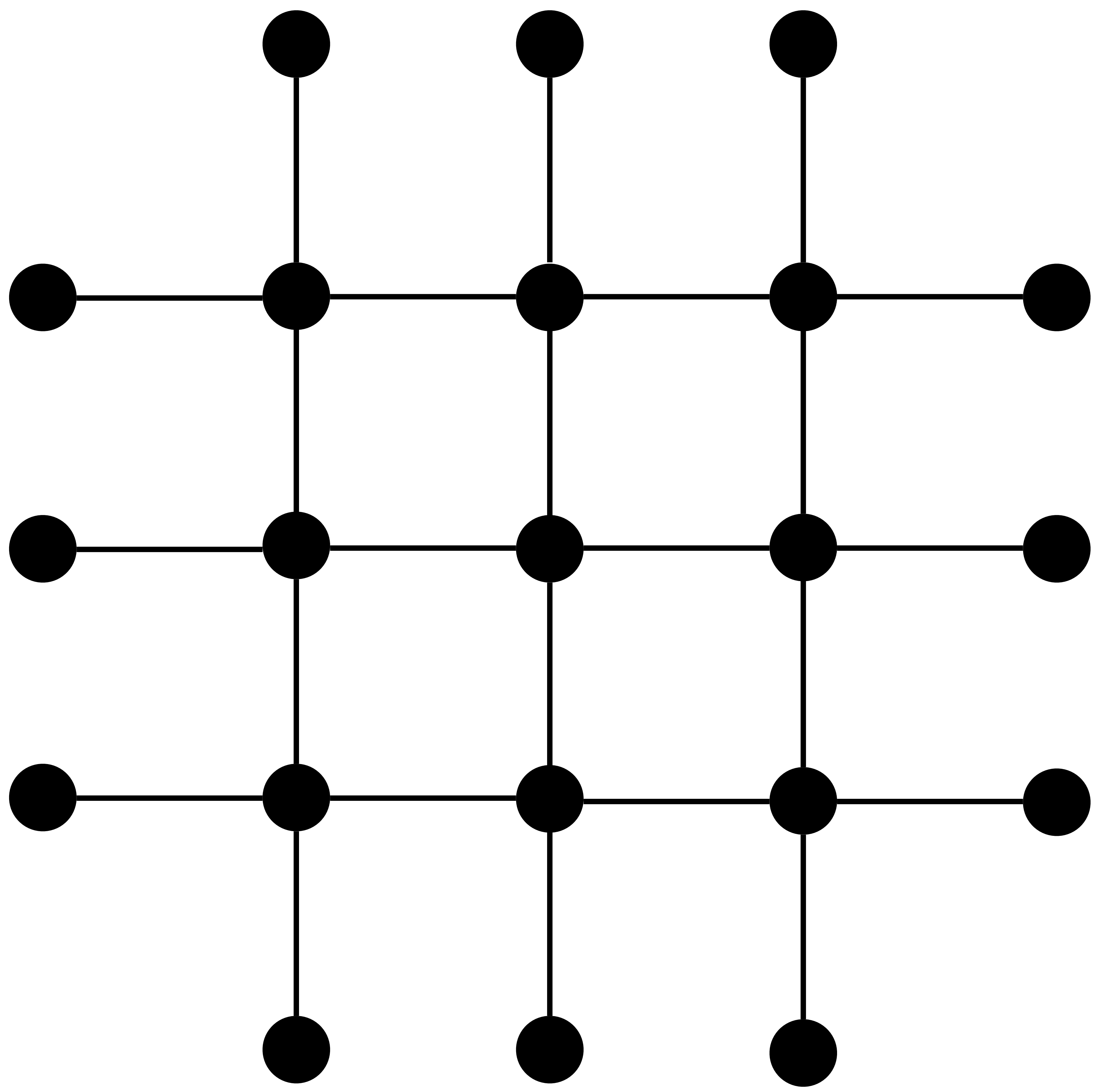
Sofic entropy and the (relative) f-invariant

Chris Shriver (UCLA) – 24 November 2020

1. Entropy on integer lattices
2. Entropy on free groups
3. Relative f-invariant

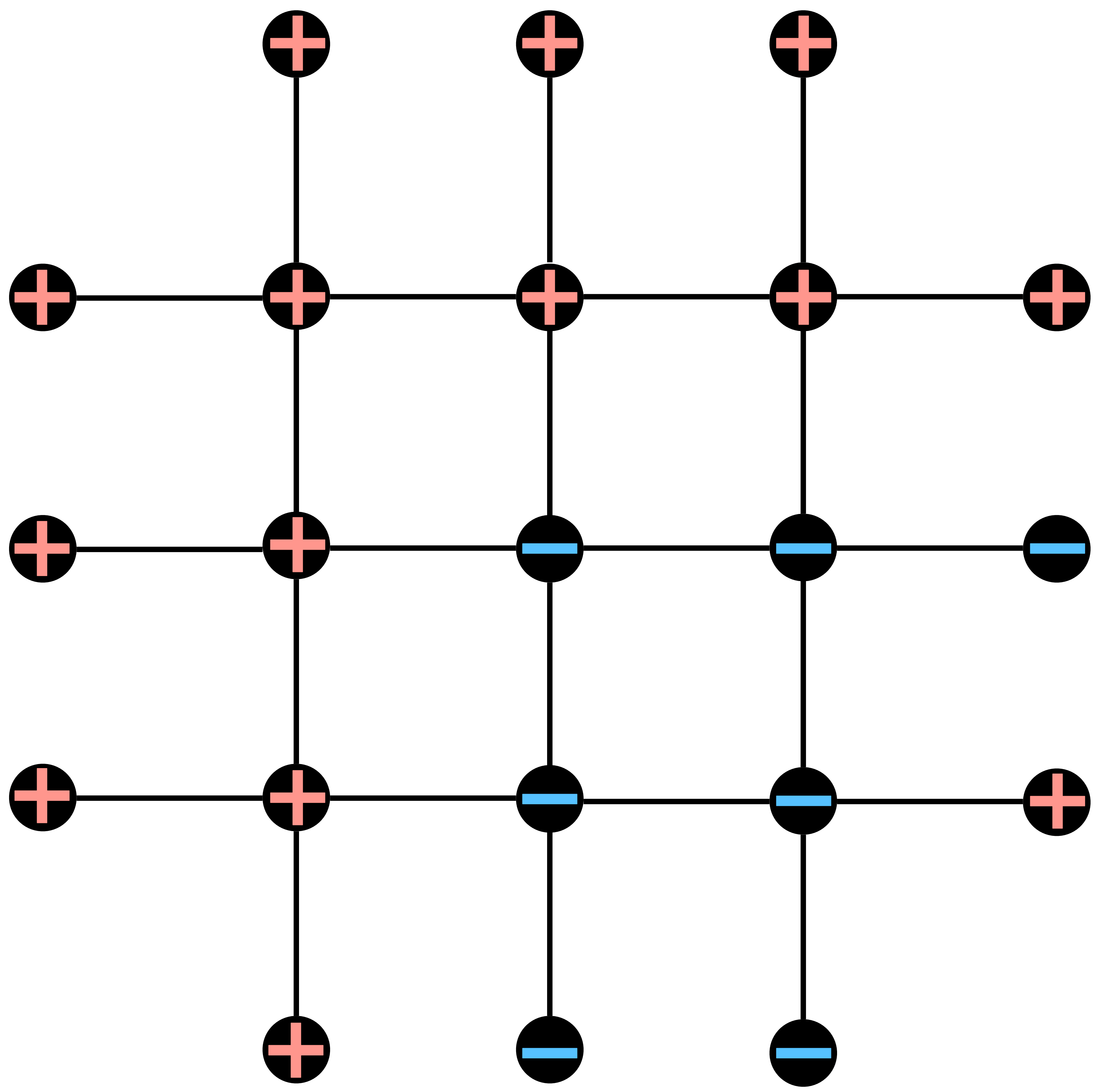
1. Entropy on integer lattices

Setup



- $\mathbb{Z}^d$ integer lattice w/ origin 0
- $\mathtt{A}$ = finite set, "alphabet"
  - ex. $\{\pm 1\}$
- $\mathbf{x} \in \mathtt{A}^{\mathbb{Z}^d}$ is a **microstate**
- $\mu \in \mathrm{Prob}(\mathtt{A}^{\mathbb{Z}^d})$ is a **state**
  - We'll assume μ shift-invariant i.e. $\left(\mathtt{A}^{\mathbb{Z}^d}, \mu, \mathbb{Z}^d\right)$ is a measure-preserving system

Setup



- $\mathbb{Z}^d$ integer lattice w/ origin 0
- $\mathtt{A}$ = finite set, "alphabet"
  - ex. $\{\pm 1\}$
- $\mathbf{x} \in \mathtt{A}^{\mathbb{Z}^d}$ is a **microstate**
- $\mu \in \mathrm{Prob}(\mathtt{A}^{\mathbb{Z}^d})$ is a **state**
  - We'll assume μ shift-invariant i.e. $\left(\mathtt{A}^{\mathbb{Z}^d}, \mu, \mathbb{Z}^d\right)$ is a measure-preserving system

Entropy rate

- The state of a system is specified by $\mu \in \mathrm{Prob}\left(\mathtt{A}^{\mathbb{Z}^d}\right)$.
  How random is μ?

- The Kolmogorov-Sinai entropy rate is defined by

Shannon entropy

$$\mathrm{h}^{\mathrm{KS}}(\mu) = \lim_{r \to \infty} \frac{1}{|B(0,r)|} \mathrm{H}\left(\mu^r\right).$$

marginal on $B(0,r)$

Lemma 1

Shannon entropy and counting

Suppose $p \in \mathrm{Prob}(\mathtt{A})$.

Then the number of $\mathbf{x} \in \mathtt{A}^n$ with

$$\frac{1}{n} \left| \{ i \in [n] : \mathbf{x}(i) = \mathtt{a} \} \right| \approx p(\{\mathtt{a}\}) \qquad \text{for all } \mathtt{a} \in \mathtt{A}$$

is about

$$\exp\big[n \, \mathrm{H}(p) \big].$$

Lemma 1

Shannon entropy and counting

Suppose $p \in \mathrm{Prob}(\mathtt{A})$.

Then the number of $\mathbf{x} \in \mathtt{A}^n$ with

$$P^0_{\mathbf{x}}(\{\mathtt{a}\}) = \frac{1}{n}\,|\,\{i \in [n] : \mathbf{x}(i) = \mathtt{a}\}\,| \approx p(\{\mathtt{a}\}) \qquad \text{for all } \mathtt{a} \in \mathtt{A}$$

is about

"empirical distribution" $\in \mathrm{Prob}(\mathtt{A})$

$$\exp\big[n\,\mathrm{H}(p)\big].$$

Lemma 1

Shannon entropy and counting

Suppose $p \in \mathrm{Prob}(\mathtt{A})$.

Then the number of $\mathbf{x} \in \mathtt{A}^n$ with

$$P^0_{\mathbf{x}}(\{\mathtt{a}\}) = \frac{1}{n}\,|\,\{i \in [n] : \mathbf{x}(i) = \mathtt{a}\}\,| \approx p(\{\mathtt{a}\}) \qquad \text{for all } \mathtt{a} \in \mathtt{A}$$

is about

“empirical distribution” $\in \mathrm{Prob}(\mathtt{A})$

$$\exp\big[n\,\mathrm{H}(p)\big].$$

Rearranged, and more precisely:

$$\mathrm{H}(p) = \lim_{\varepsilon \downarrow 0} \limsup_{n \to \infty} \frac{1}{n} \log |\,\{\mathbf{x} \in \mathtt{A}^n : \|P^0_{\mathbf{x}} - p\|_{TV} < \varepsilon\}\,|$$

Entropy rate via counting

$$\mathrm{H}(p) = \lim_{\varepsilon \downarrow 0} \limsup_{n \to \infty} \frac{1}{n} \log |\{\mathbf{x} \in \mathtt{A}^n : \|P^0_{\mathbf{x}} - p\|_{TV} < \varepsilon\}|$$

The KS entropy can be expressed in a similar form:

$$\mathrm{h}^{\mathrm{KS}}(\mu) = \inf_{r,\varepsilon} \limsup_{n \to \infty} \frac{1}{|B(0,n)|} \log |\{\mathbf{x} \in \mathtt{A}^{B(0,n)} : \|P^r_{\mathbf{x}} - \mu^r\|_{TV} < \varepsilon\}|$$

where

$\mathbf{x}$ is a “microstate on a finite subsystem”

$P^r_{\mathbf{x}}$ is a “radius-r empirical distribution” to be defined.

- This is a special case of the fact that (sofic entropy) = (KS entropy) for amenable groups, proven by Bowen [2010]

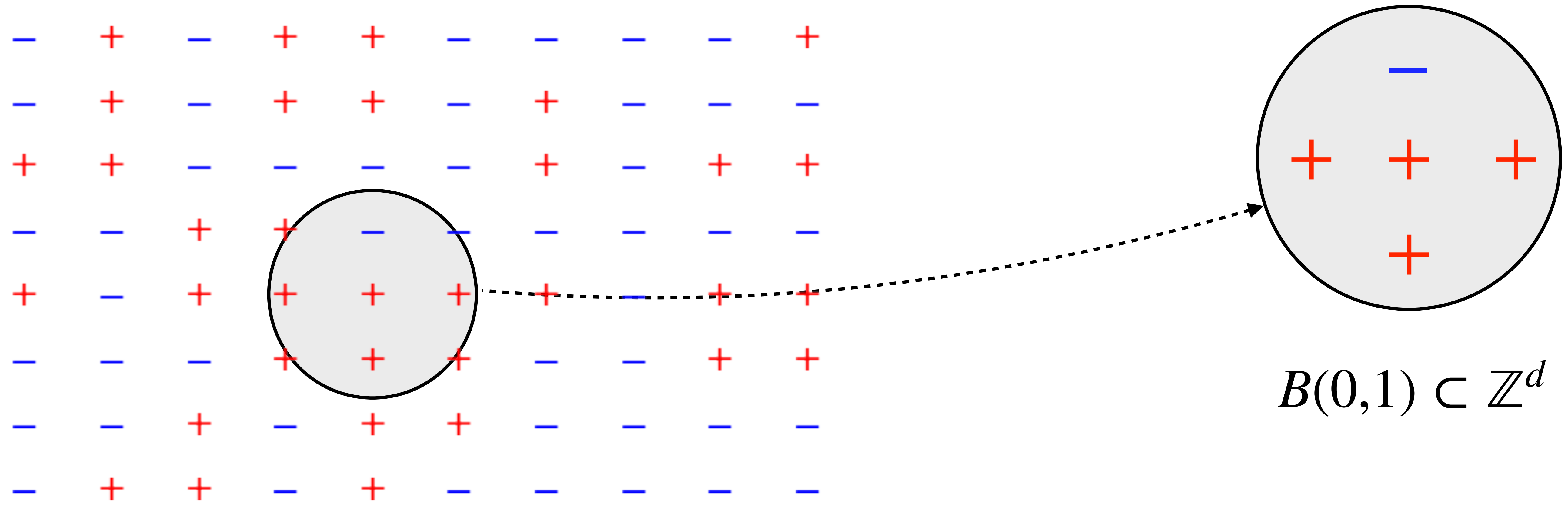
Local statistics of microstates

Let $K \Subset \mathbb{Z}^d$ be a large rectangle.

Given a microstate $\mathbf{x} \in \mathtt{A}^K$ and a radius $r \in \mathbb{N}$, the depth-r **neighborhood labeling** $\mathbf{x}^r$ is the element of $\left(\mathtt{A}^{B(0,r)}\right)^K$ given by

$$\mathbf{x}^r(v) \doteq \big(\mathbf{x}(v+w)\big)_{w \in B(0,r)}.$$

$\mathtt{A} = \{\, +\, ,\, -\, \}$



$K \subset \mathbb{Z}^d$

$B(0,1) \subset \mathbb{Z}^d$

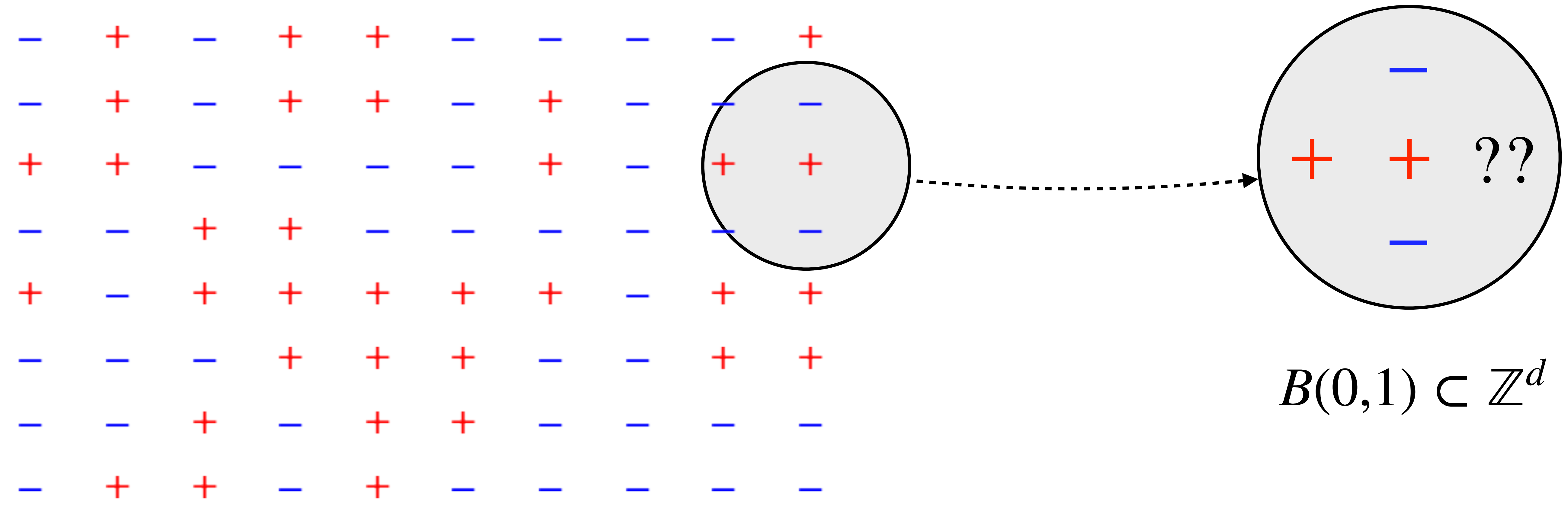
Local statistics of microstates

Let $K \Subset \mathbb{Z}^d$ be a large rectangle.

Given a microstate $\mathbf{x} \in \mathtt{A}^K$ and a radius $r \in \mathbb{N}$, the depth-r **neighborhood labeling** $\mathbf{x}^r$ is the element of $\left(\mathtt{A}^{B(0,r)}\right)^K$ given by

$$\mathbf{x}^r(v) \doteq \big(\mathbf{x}(v+w)\big)_{w \in B(0,r)}.$$

$\mathtt{A} = \{ +, - \}$



$B(0,1) \subset \mathbb{Z}^d$

$K \subset \mathbb{Z}^d$

Empirical distribution and good models

- The radius-r **empirical distribution** of $\mathbf{x} \in \mathtt{A}^K$ is defined by

$$P_{\mathbf{x}}^r = P_{\mathbf{x}^r}^0 \in \operatorname{Prob}\left(\mathtt{A}^{B(0,r)}\right).$$

Empirical distribution and good models

- The radius-r **empirical distribution** of $\mathbf{x} \in \mathtt{A}^K$ is defined by

$$P^r_{\mathbf{x}} = P^0_{\mathbf{x}^r} \in \operatorname{Prob}\left(\mathtt{A}^{B(0,r)}\right).$$

- Say $\mathbf{x}$ is an (r, ε)**-good model** for $\mu \in \operatorname{Prob}\left(\mathtt{A}^{\mathbb{Z}^d}\right)$ if

$$\|P^r_{\mathbf{x}} - \mu^r\| < \varepsilon.$$

- $\Omega(K, \mu, r, \varepsilon)$ is the set of such $\mathbf{x} \in \mathtt{A}^K$.

Summary

and some terminology

With $\Sigma = (K_n)_{n=1}^{\infty}$ a (Følner) sequence of boxes which exhausts $\mathbb{Z}^d$, we have

$$\mathrm{h}_{\Sigma}(\mu) = \inf_{r,\varepsilon} \limsup_{n\to\infty} \frac{1}{n} \log |\Omega(K_n, \mu, r, \varepsilon)| = \mathrm{h}^{KS}(\mu)$$

Σ is a sofic approximation to $\mathbb{Z}^d$ because

1. we have approximately free approximate actions of $\mathbb{Z}^d$ on K_n
2. K_n locally look like $\mathbb{Z}^d$

2. Entropy for free-group actions

Based on:
Lewis Bowen. “The Ergodic Theory of Free Group Actions: Entropy and the f-Invariant.” *Groups, Geometry, and Dynamics* (2010), pp. 419-432.

Free group setup

$\mathbb{F}_r = \langle s_1, \ldots, s_r \rangle$ = rank-r free group with identity e

Let $\mathbf{z}$ be a $\mathtt{A}^{\mathbb{F}_r}$-valued r.v. with shift-invariant law μ. The **f-invariant** of μ is defined by

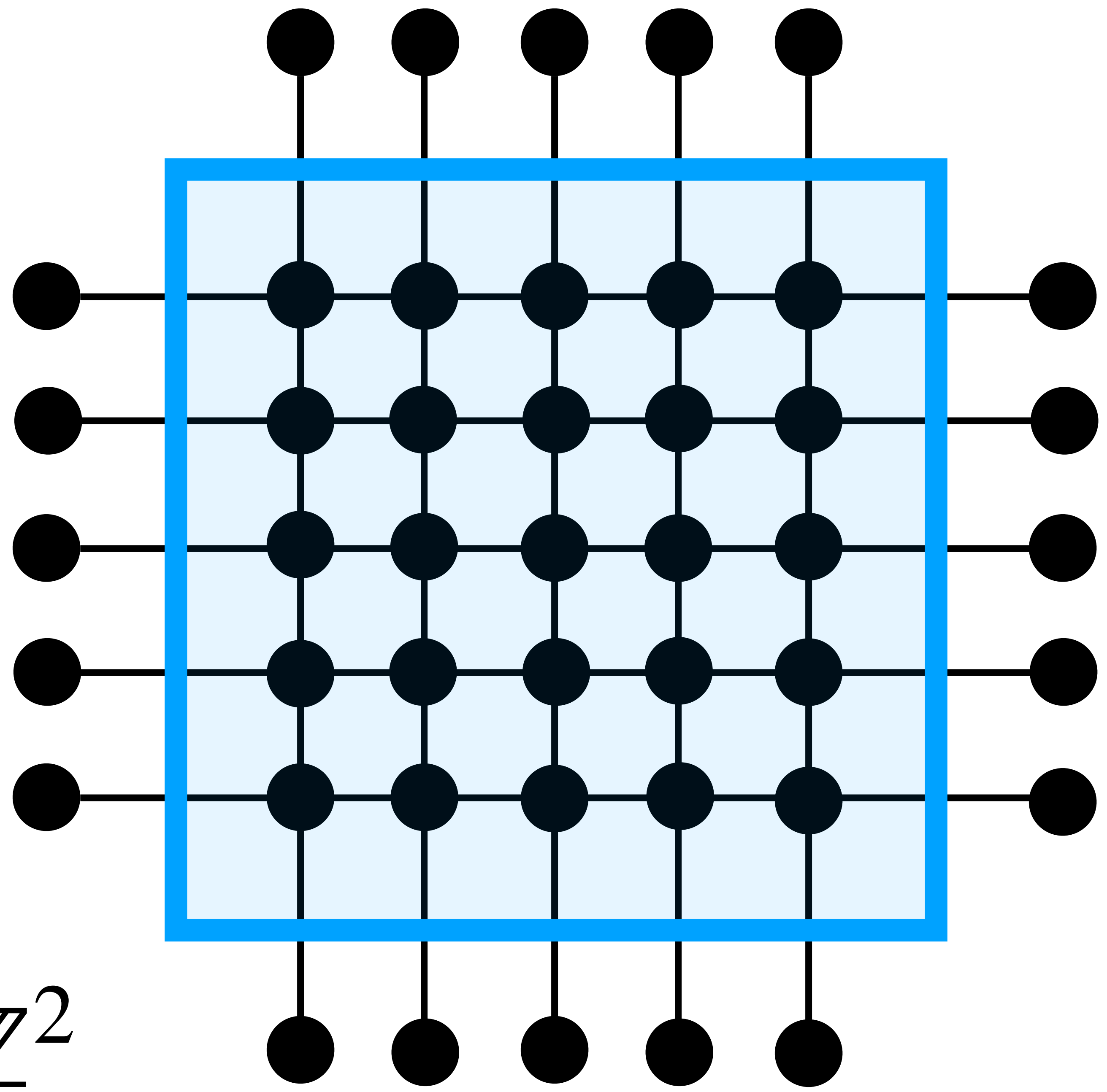
$$f(\mu) = \inf_R \left(\mathrm{H}\big(\mathbf{z}^R(e)\big) - \sum_{i=1}^{r} \mathrm{I}\big(\mathbf{z}^R(e); \mathbf{z}^R(s_i)\big) \right).$$

where $\mathrm{I}(\mathbf{x};\mathbf{y}) = \mathrm{H}(\mathbf{x}) + \mathrm{H}(\mathbf{y}) - \mathrm{H}(\mathbf{x},\mathbf{y})$ & $\mathbf{z}^R$ is a $(\mathtt{A}^{B(e,R)})^{\mathbb{F}_r}$-valued r.v. with law μ^R

Is there a “counting good models” formula for $f(\mu)$?

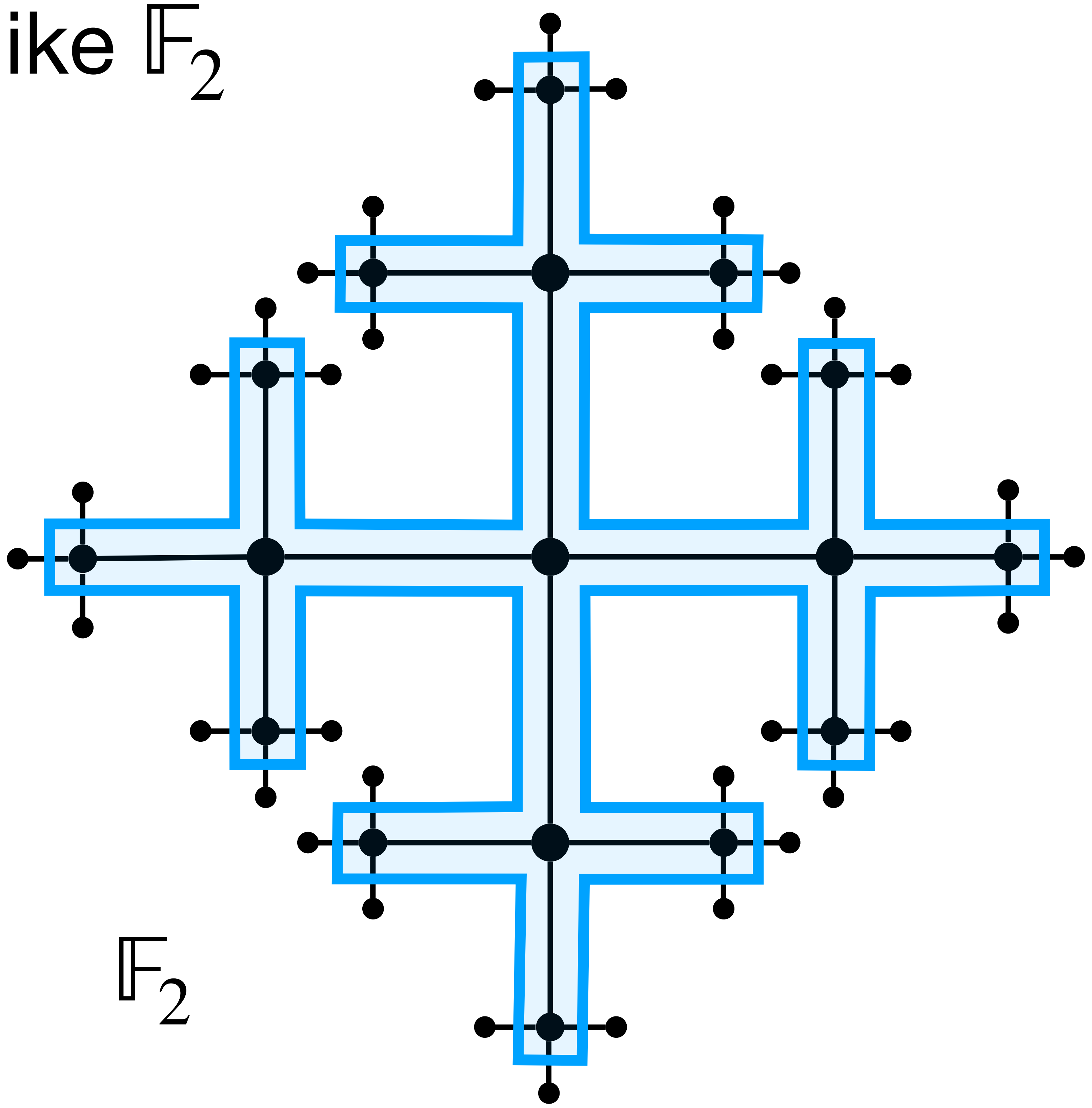
Microstates for free groups

- For $\mathbf{x} \in \mathtt{A}^{B(e,R)}$ we can't make sense of the empirical distribution because most sites are close to the edge.
- A large finite subgraph doesn't locally look like $\mathbb{F}_2$



$\mathbb{Z}^2$

vs.



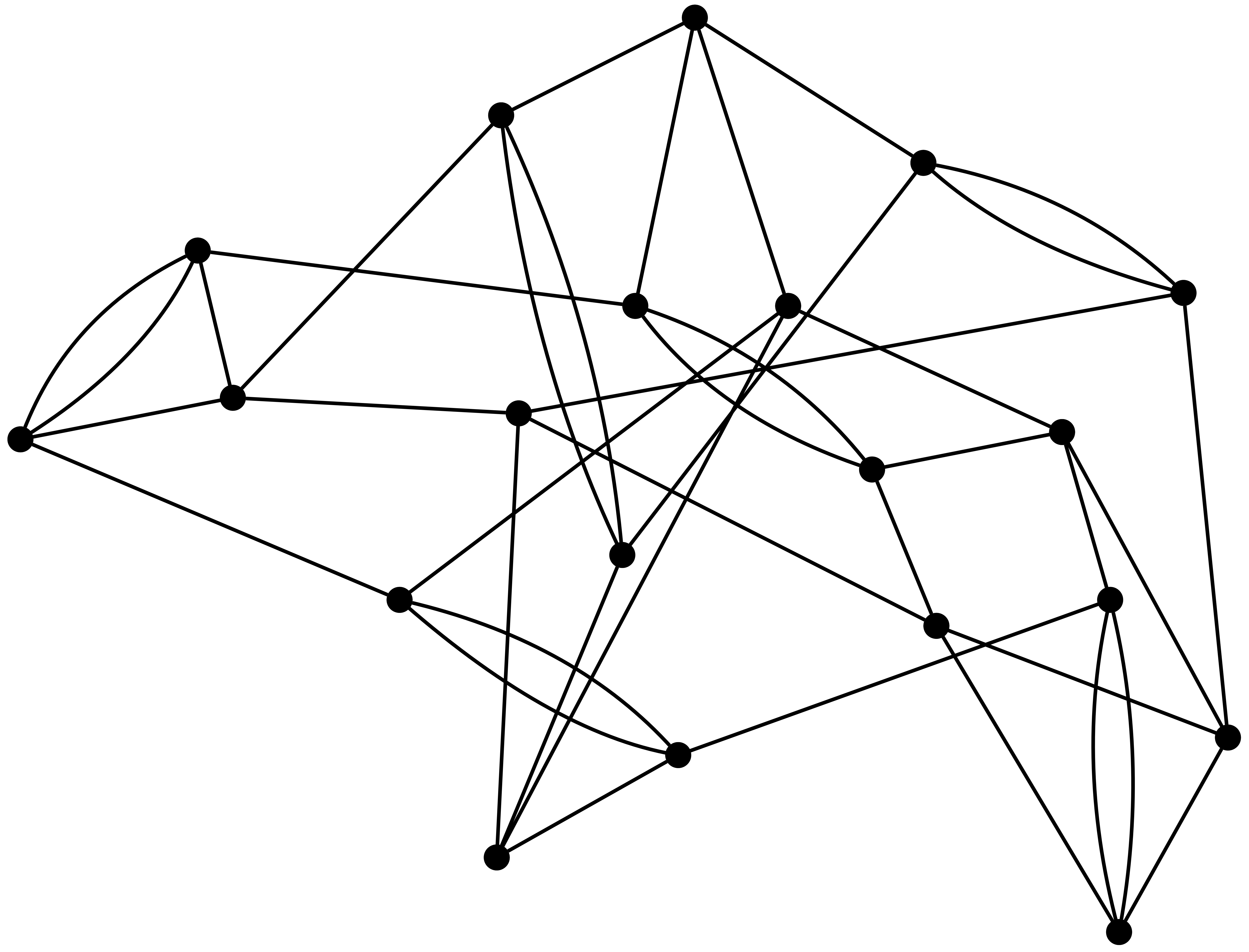
$\mathbb{F}_2$

Partial fix

- A random regular graph G "locally looks like $\mathbb{F}_r$" in that:

  For any R, the fraction of vertices $v \in G$ with $B^G(v, R) \cong B^\Gamma(e, R)$ converges in prob. to 1 as size of $G \to \infty$.

- But how do we make sense of $\mathbf{x}^R$ and $P_{\mathbf{x}}^R$?

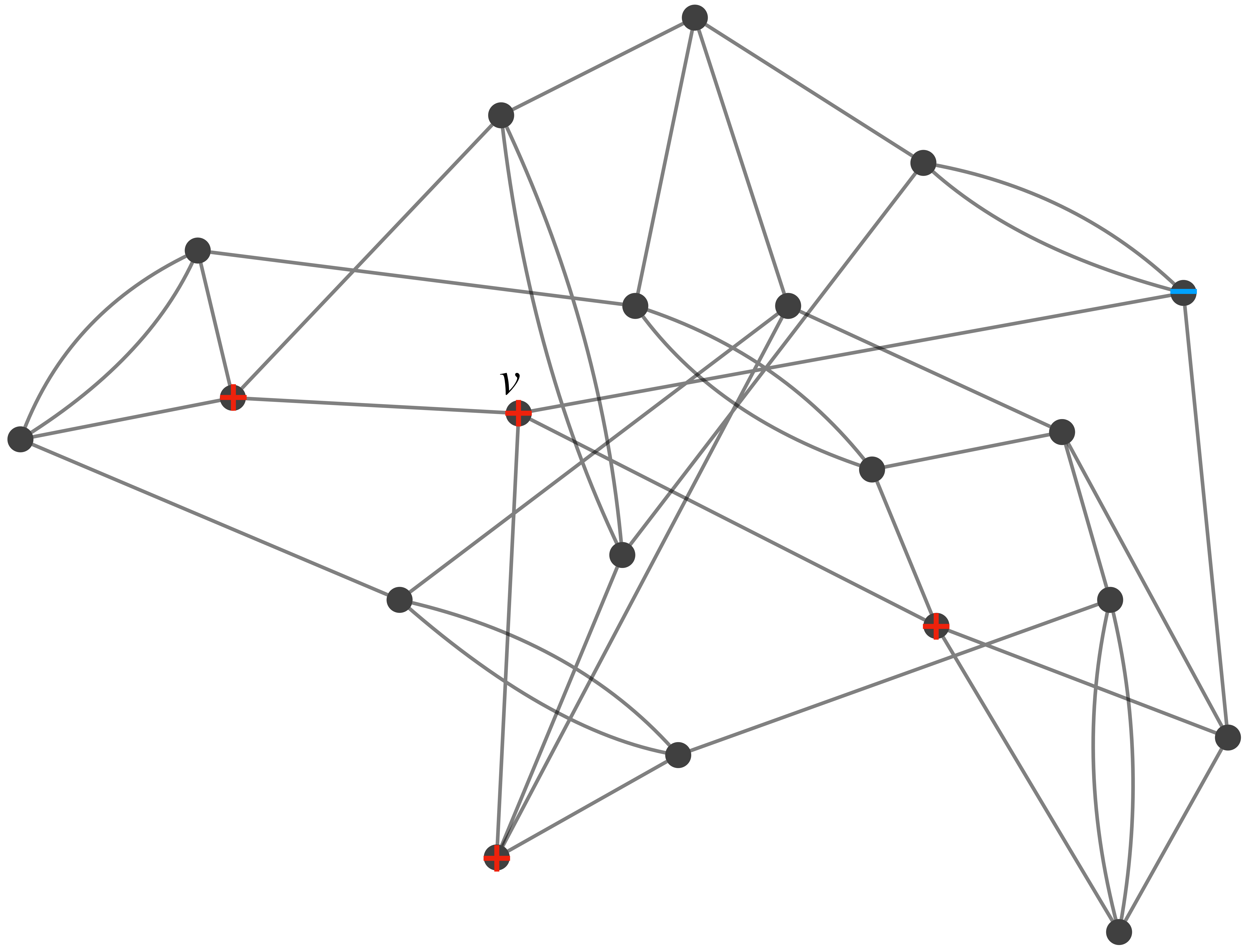


Partial fix

- A random regular graph G "locally looks like $\mathbb{F}_r$" in that:

  For any R, the fraction of vertices $v \in G$ with $B^G(v, R) \cong B^\Gamma(e, R)$ converges in prob. to 1 as size of $G \to \infty$.

- But how do we make sense of $\mathbf{x}^R$ and $P^R_\mathbf{x}$?

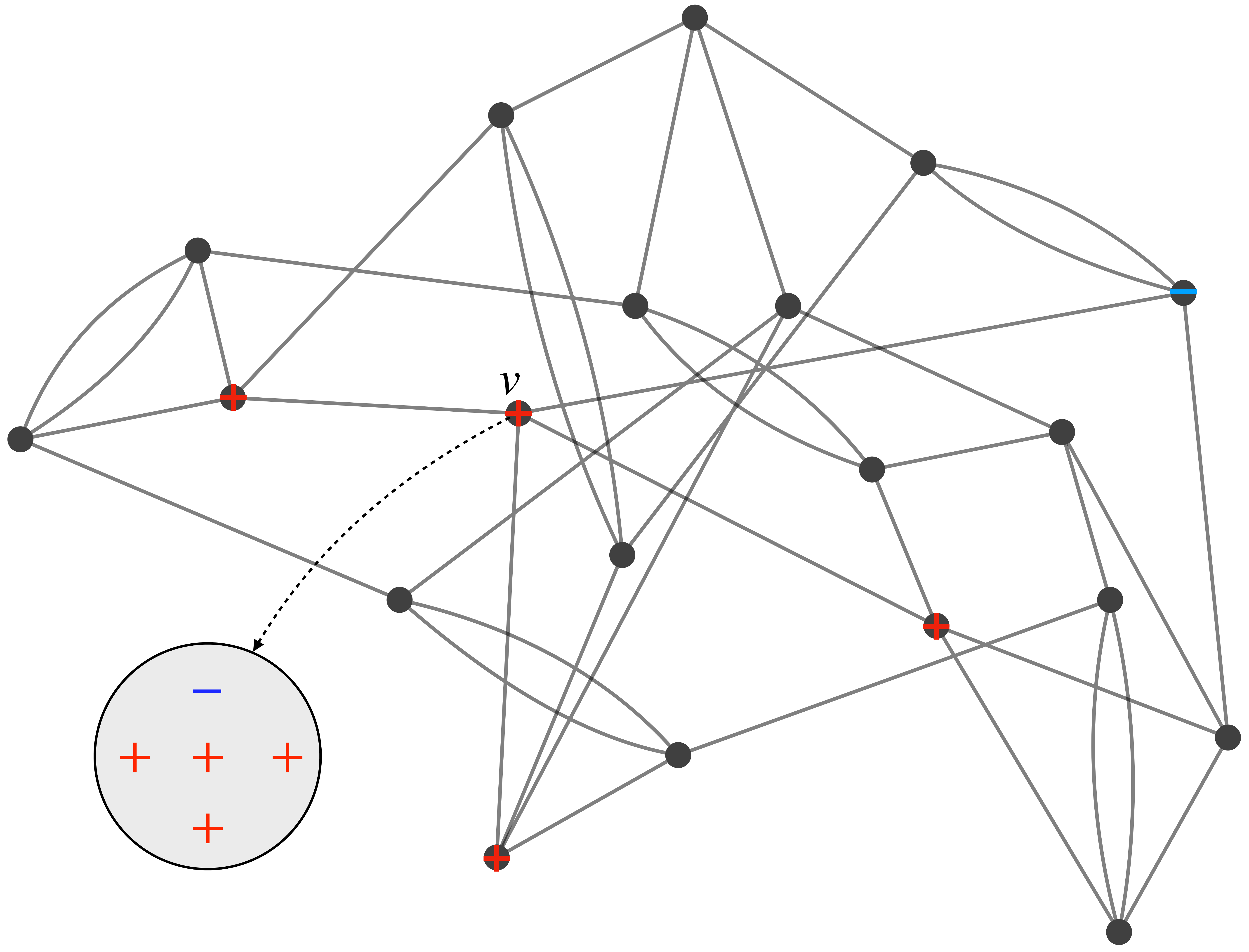


Partial fix

- A random regular graph G "locally looks like $\mathbb{F}_r$" in that:

  For any R, the fraction of vertices $v \in G$ with $B^G(v, R) \cong B^\Gamma(e, R)$ converges in prob. to 1 as size of $G \to \infty$.

- But how do we make sense of $\mathbf{x}^R$ and $P^R_{\mathbf{x}}$?



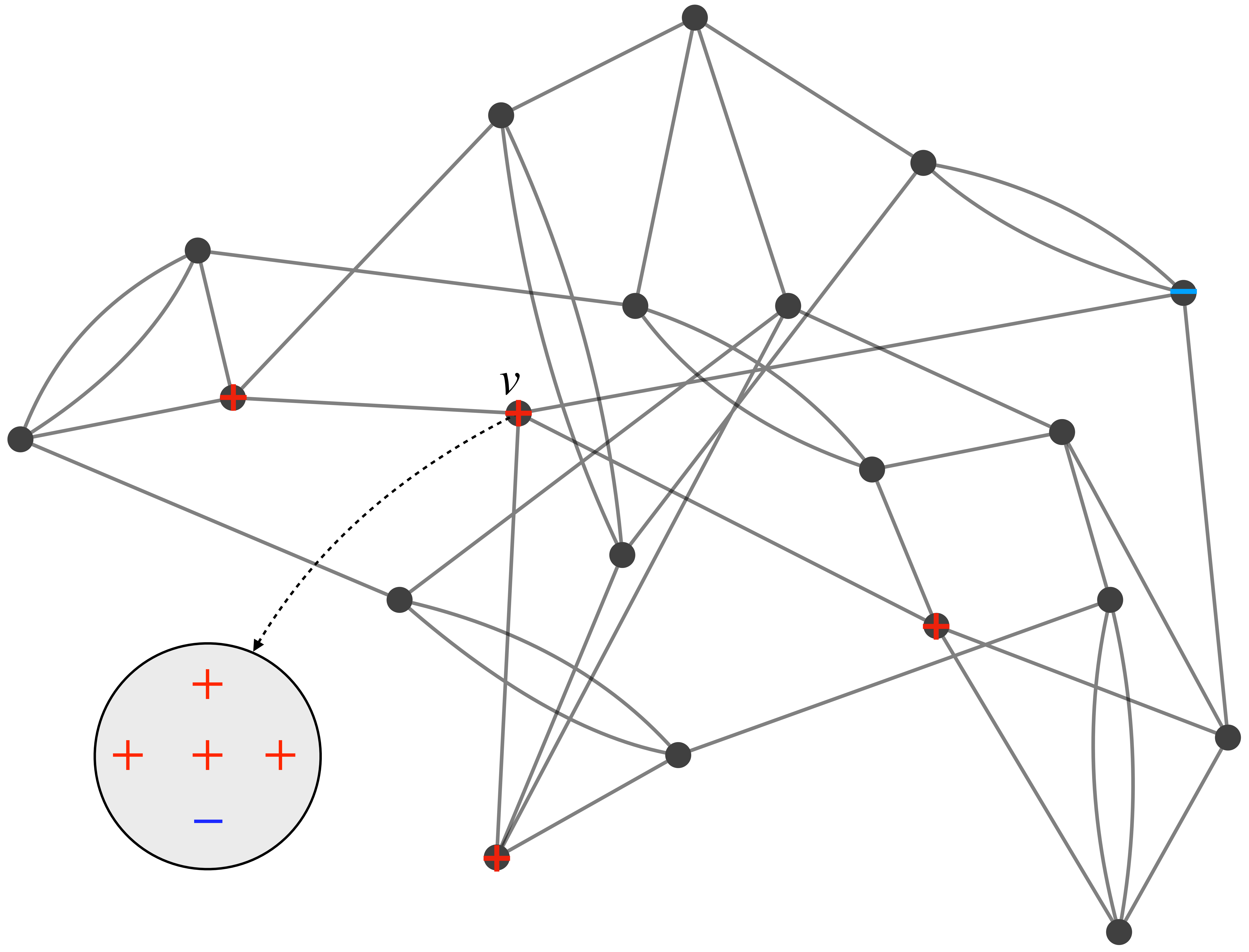
Partial fix

- A random regular graph G "locally looks like $\mathbb{F}_r$" in that:

  For any R, the fraction of vertices $v \in G$ with $B^G(v,R) \cong B^\Gamma(e,R)$ converges in prob. to 1 as size of $G \to \infty$.

- But how do we make sense of $\mathbf{x}^R$ and $P^R_{\mathbf{x}}$?

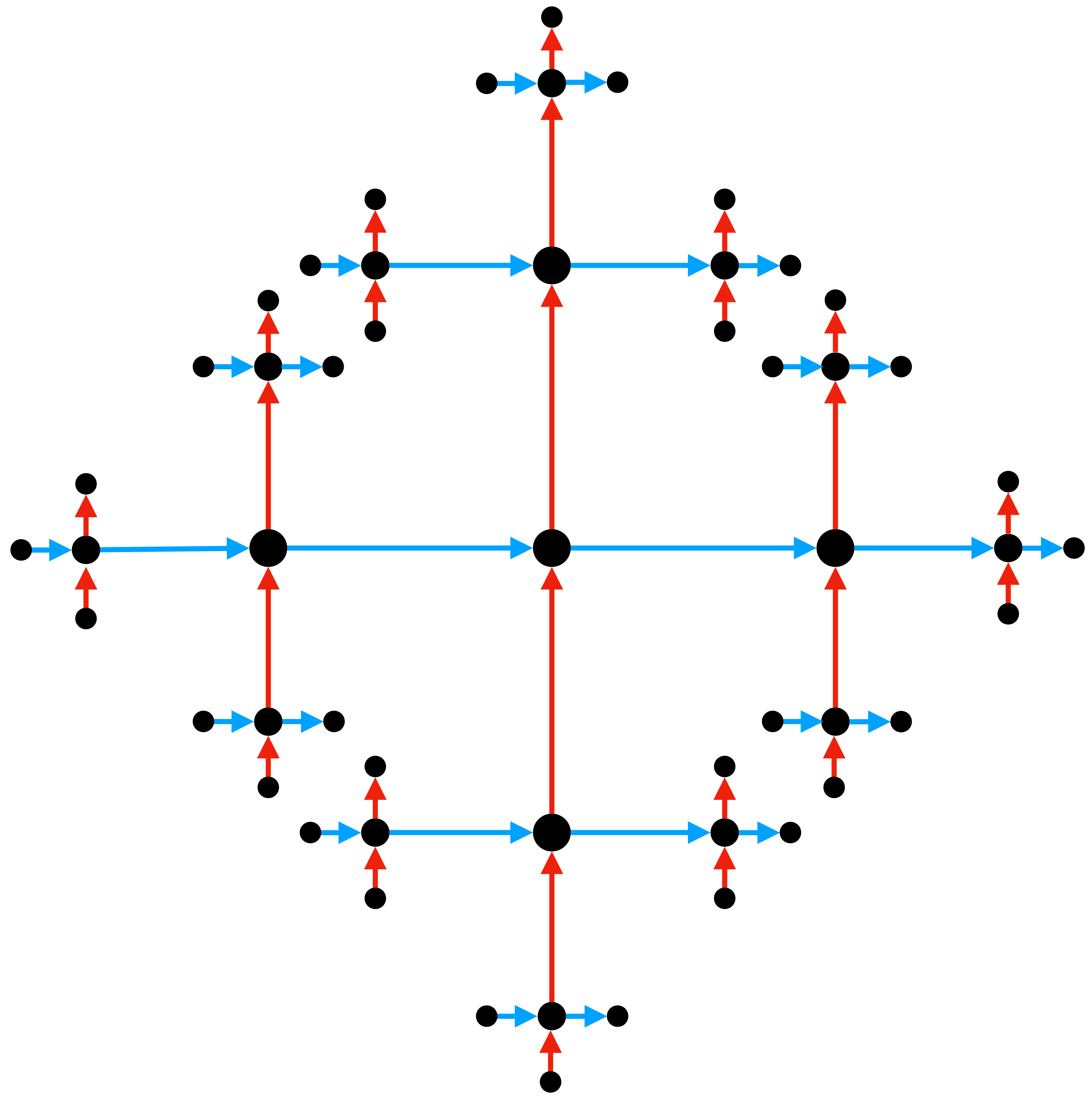
no canonical choice!



Extra graph data

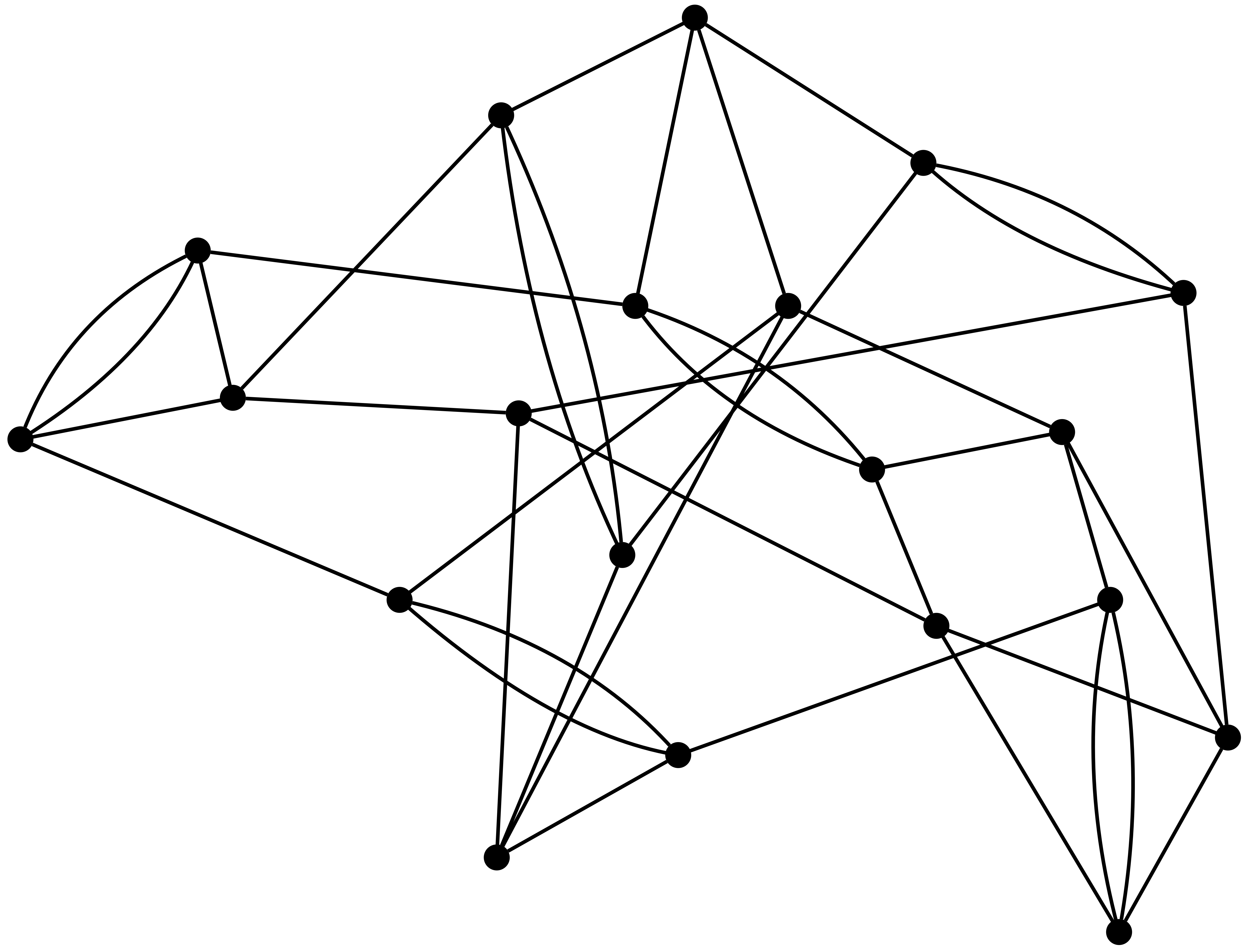
infinite systems

Edges of a Cayley graph naturally come directed and labeled by generators.



Extra graph data

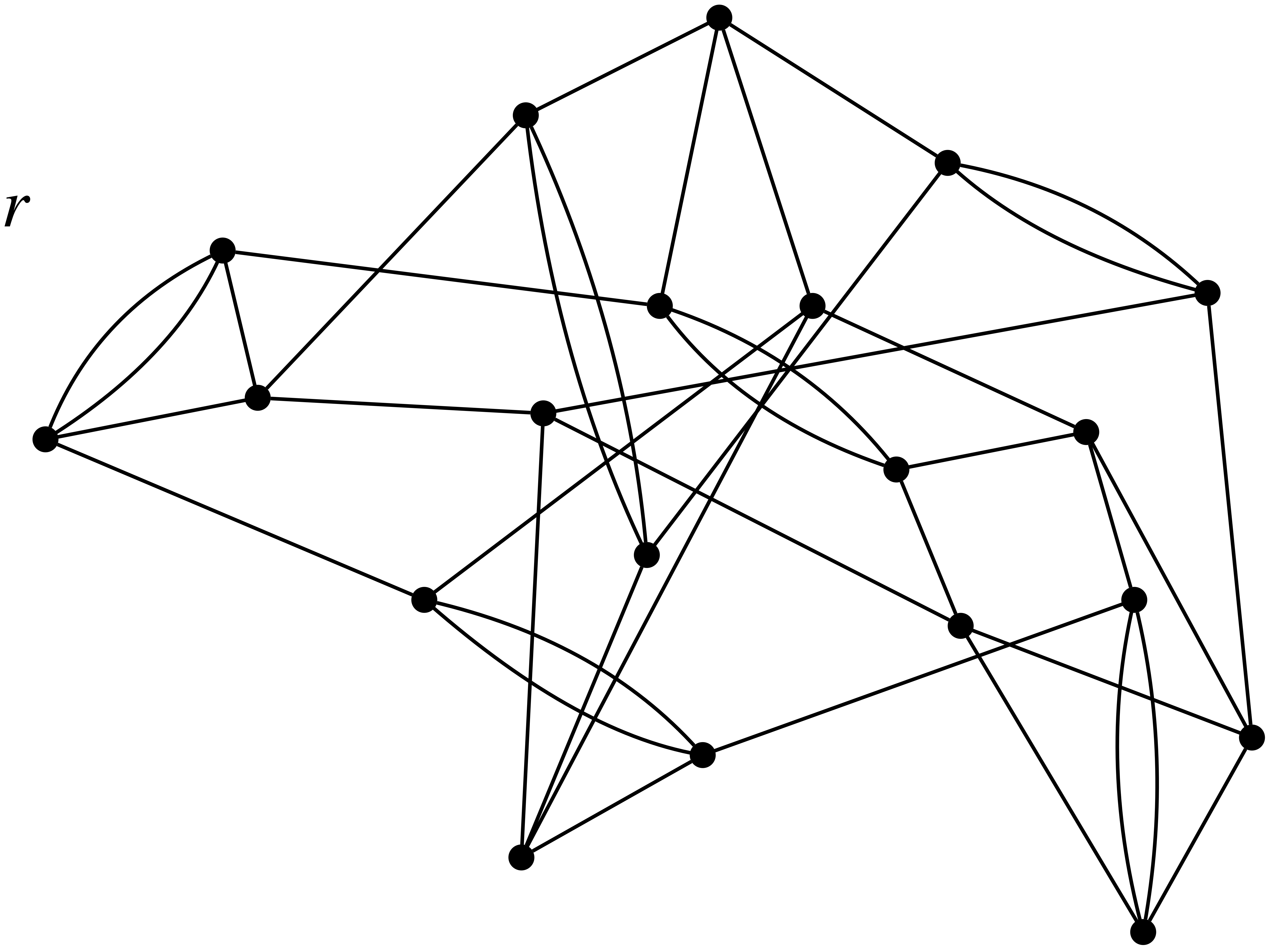
finite systems



Extra graph data

finite systems

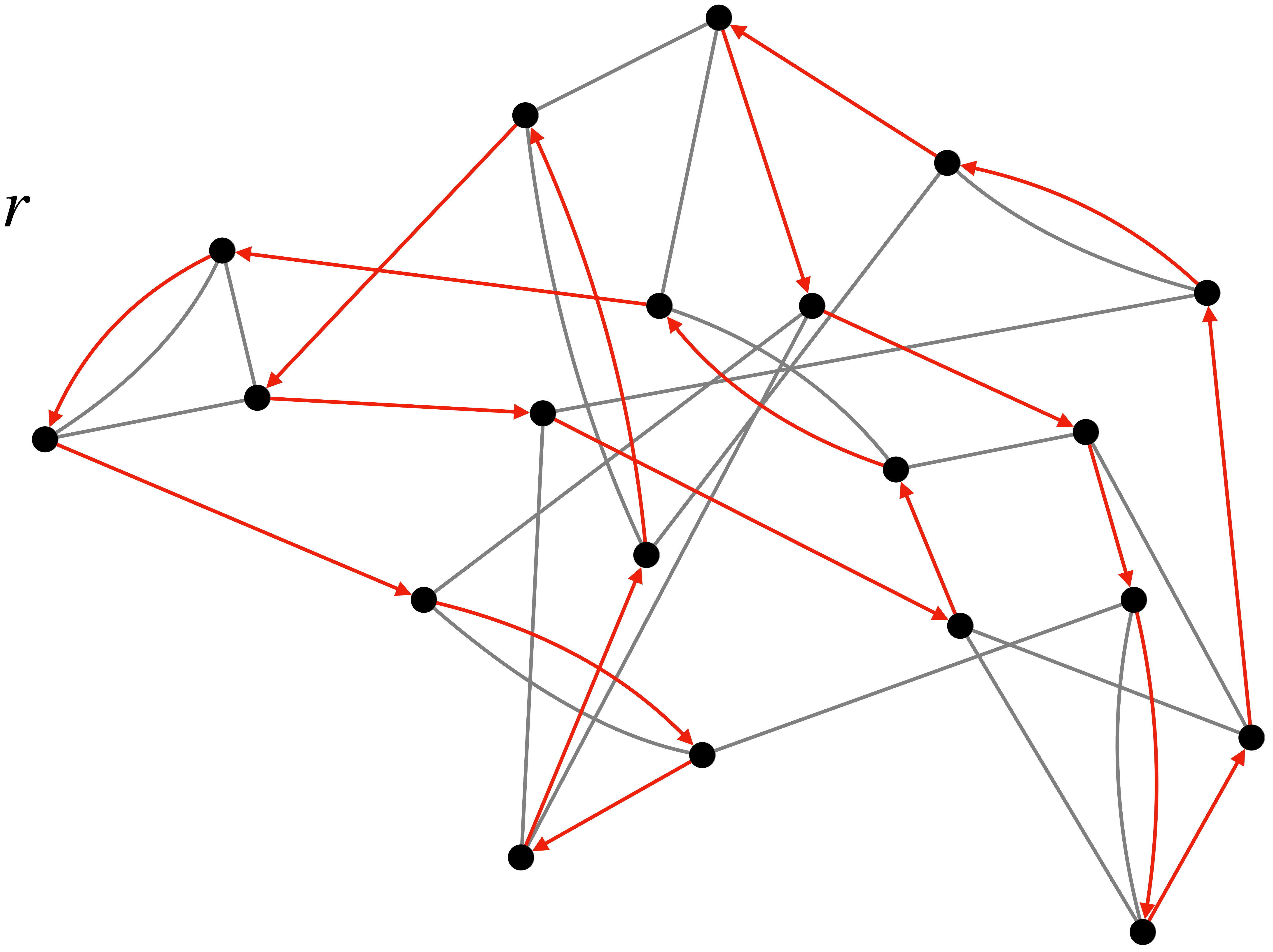
- Write $2r$-regular G as union of r permutations, and label edges



Extra graph data

finite systems

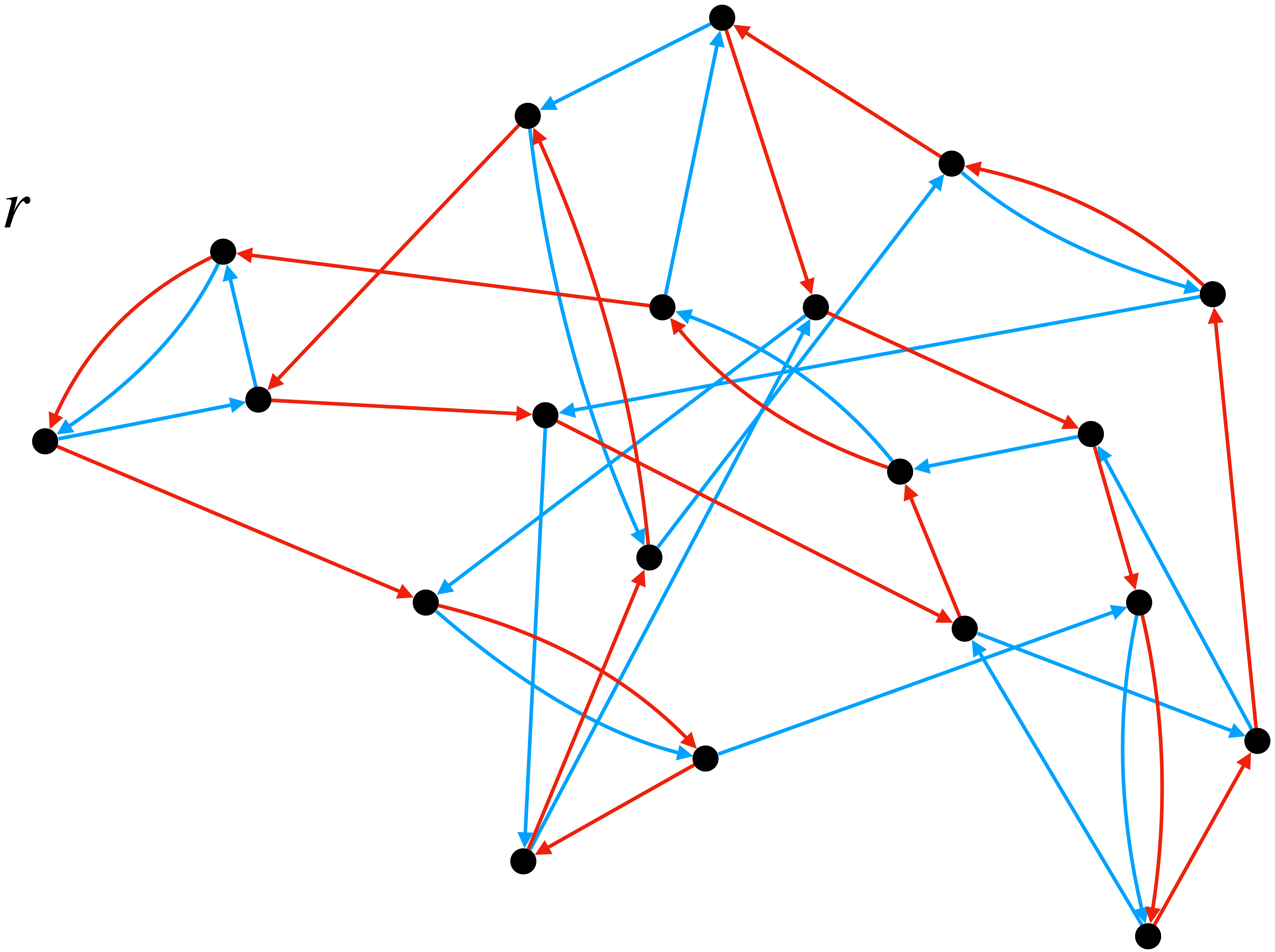
- Write $2r$-regular G as union of r permutations, and label edges



Extra graph data

finite systems

- Write $2r$-regular G as union of r permutations, and label edges

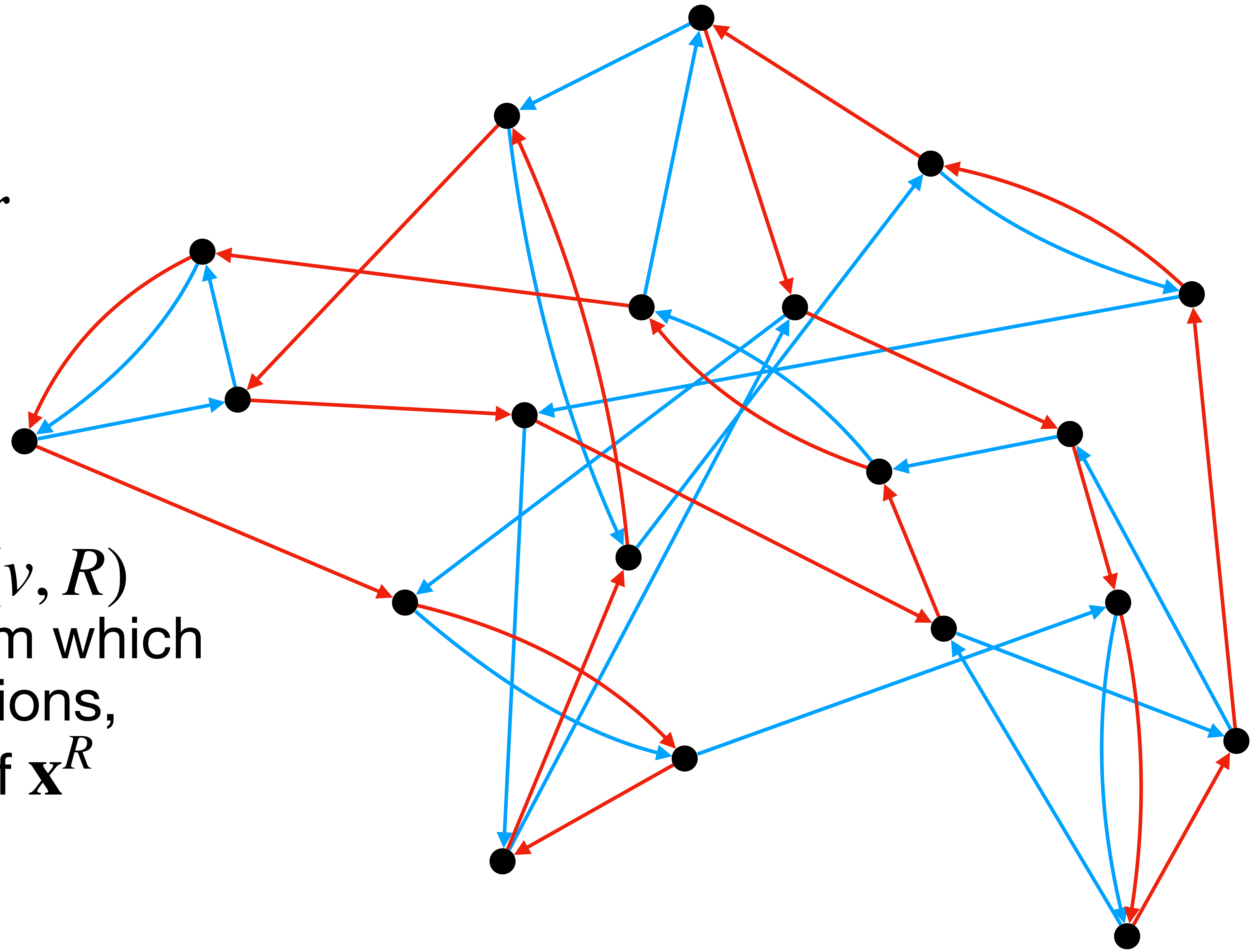


Extra graph data

finite systems

- Write $2r$-regular G as union of r permutations, and label edges

- Now whenever $\mathrm{B}^{\Gamma}(e, R) \cong \mathrm{B}^{G}(v, R)$ there is exactly one isomorphism which respects edge labels and directions, which gives a canonical defn. of $\mathbf{x}^R$

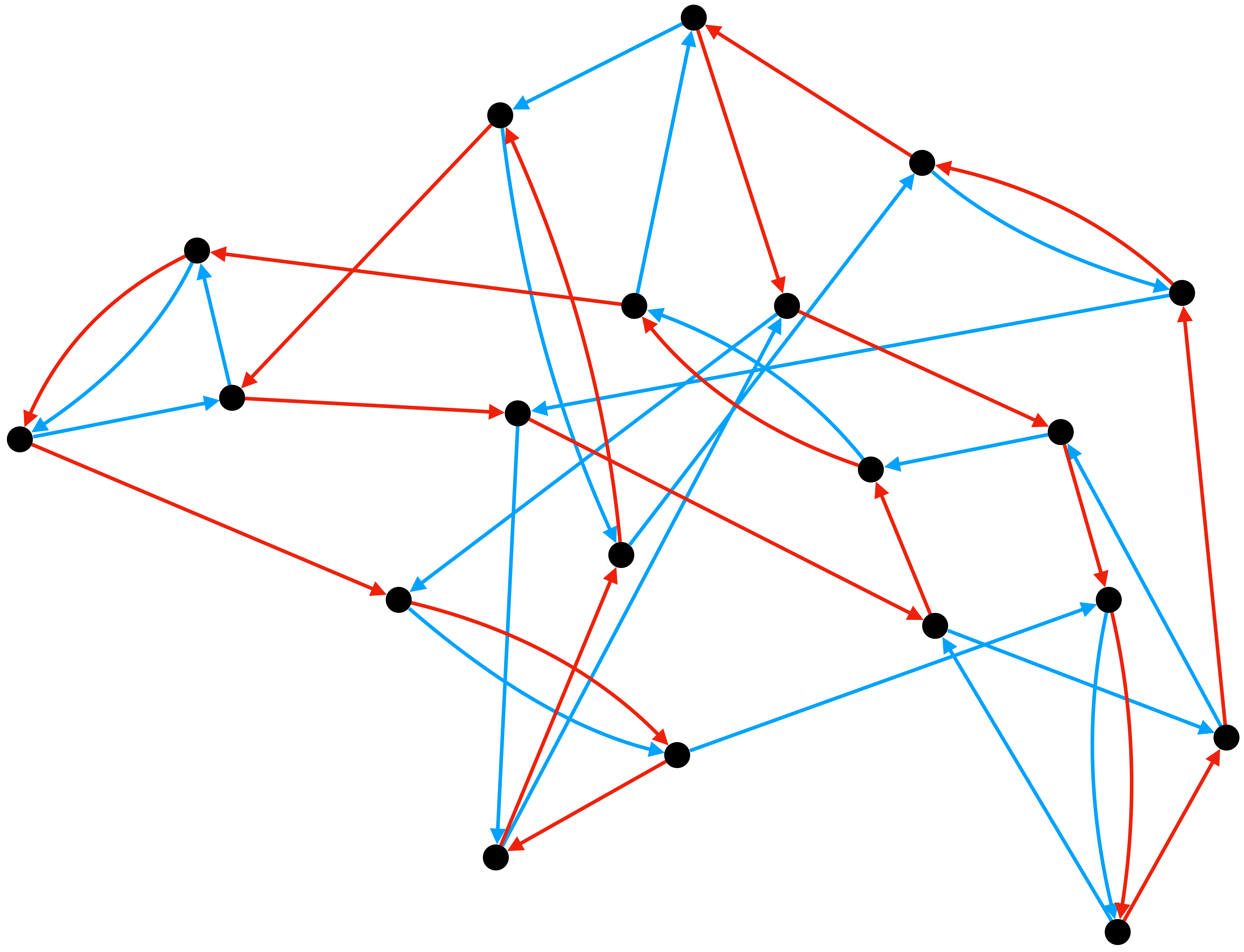


Permutation model

- Pick a random regular graph with vertex set $[n] = \{1,\ldots,n\}$ by picking r permutations $\sigma_1, \ldots, \sigma_r$ uniformly at random.

- Write $\sigma = (\sigma_1, \ldots, \sigma_r)$. Can be thought of as a random homomorphism $\mathbb{F}_r \to \mathrm{Sym}(n)$

- Let ζ_n be the law of σ (r is implicit)



The f-invariant via counting good models

Theorem [Bowen '10]

$$f(\mu) = \inf_{R,\varepsilon} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}_{\sigma\sim\zeta_n} |\Omega(\sigma,\mu,R,\varepsilon)|$$

In other words

$$f(\mu) = \mathrm{h}_{\boldsymbol{\Sigma}}(\mu)$$

where $\boldsymbol{\Sigma} = (\zeta_n)_{n=1}^{\infty}$ is a random sofic approximation.

Idea of proof

first attempt

- Want to estimate $|\Omega(\sigma, \mu, R, \varepsilon)|$,
  i.e. the number of $\mathbf{x} \in \mathtt{A}^n$ with $\|P^{\sigma,R}_{\mathbf{x}} - \mu^R\|_{\mathrm{TV}} < \varepsilon$.
- For any such $\mathbf{x}$, by definition $\mathbf{x}^R \in (\mathtt{A}^{B(e,R)})^n$ satisfies $\|P^{\sigma,0}_{\mathbf{x}^R} - \mu^R\|_{\mathrm{TV}} < \varepsilon$

Idea of proof

first attempt

- Want to estimate $|\Omega(\sigma, \mu, R, \varepsilon)|$,
  i.e. the number of $\mathbf{x} \in \mathtt{A}^n$ with $\|P_{\mathbf{x}}^{\sigma,R} - \mu^R\|_{\mathrm{TV}} < \varepsilon$.
- For any such $\mathbf{x}$, by definition $\mathbf{x}^R \in (\mathtt{A}^{B(e,R)})^n$ satisfies $\|P_{\mathbf{x}^R}^{\sigma,0} - \mu^R\|_{\mathrm{TV}} < \varepsilon$
- By Lemma 1, the number of $\mathbf{X} \in (\mathtt{A}^{B(e,R)})^n$ with this property is about $\exp\left[n\,\mathrm{H}\left(\mu^R\right)\right]$. So

Idea of proof

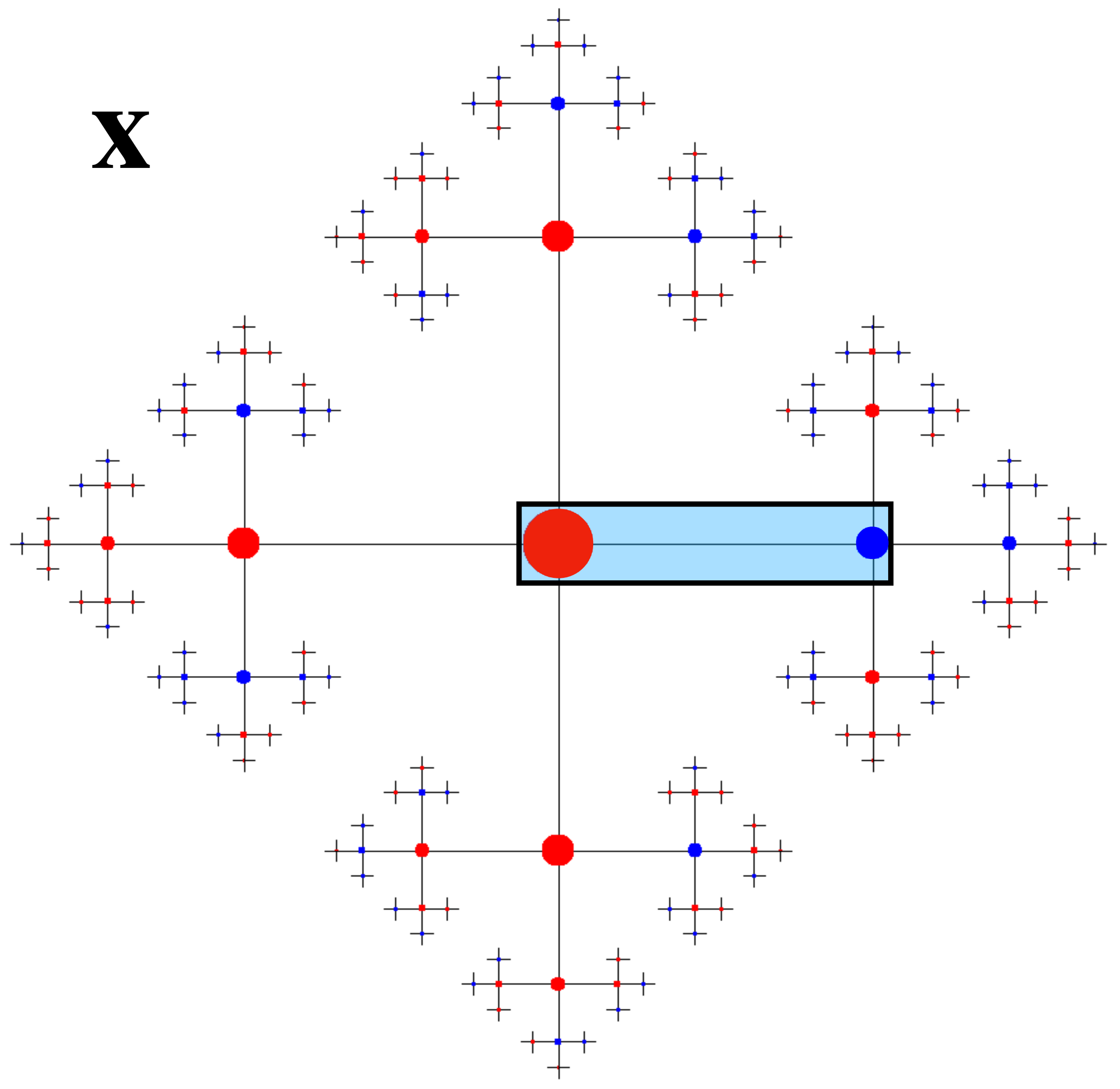
first attempt

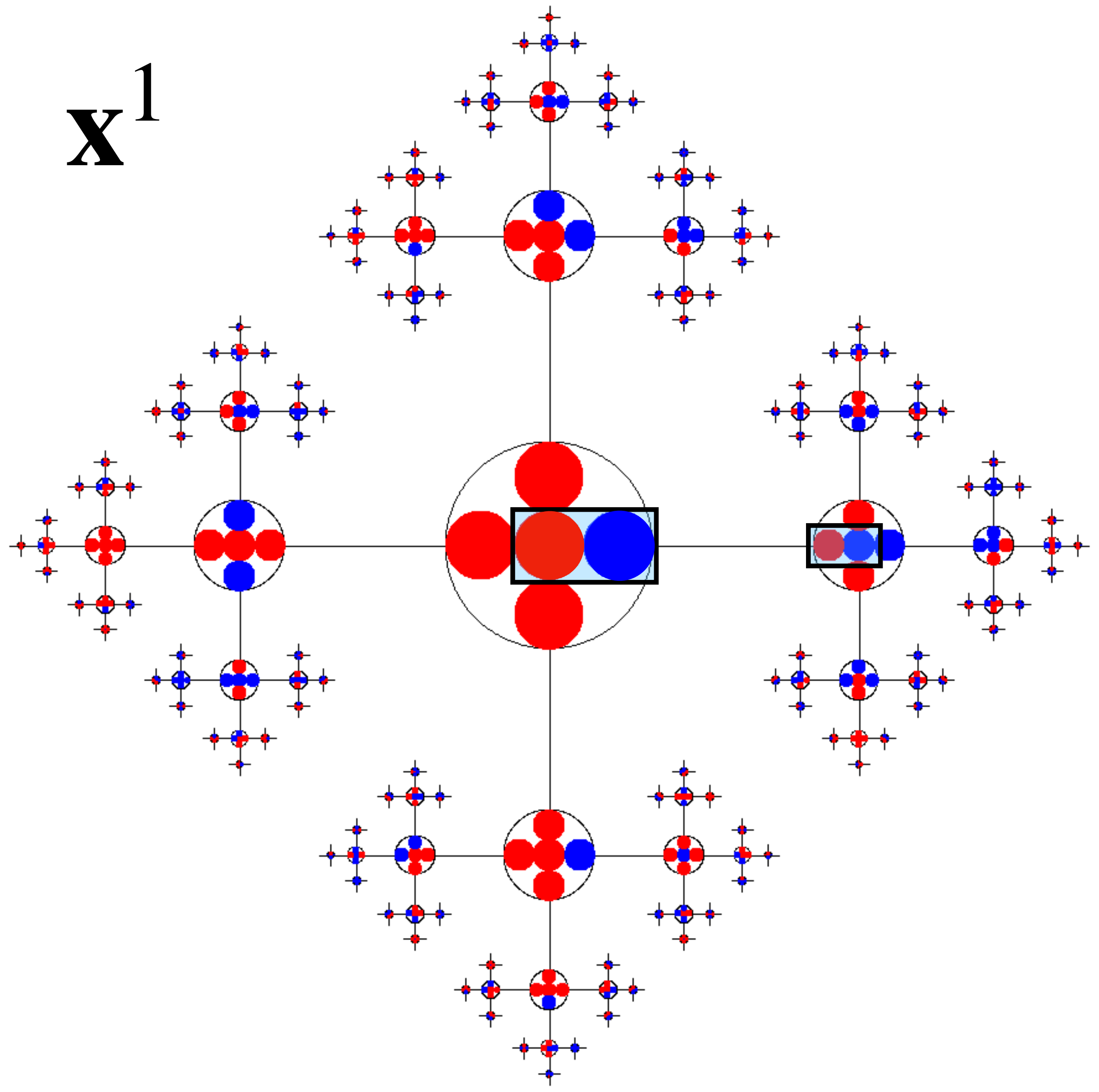
- Want to estimate $|\Omega(\sigma, \mu, R, \varepsilon)|$,
  i.e. the number of $\mathbf{x} \in \mathtt{A}^n$ with $\|P^{\sigma,R}_{\mathbf{x}} - \mu^R\|_{\mathrm{TV}} < \varepsilon$.
- For any such $\mathbf{x}$, by definition $\mathbf{x}^R \in (\mathtt{A}^{B(e,R)})^n$ satisfies $\|P^{\sigma,0}_{\mathbf{x}^R} - \mu^R\|_{\mathrm{TV}} < \varepsilon$
- By Lemma 1, the number of $\mathbf{X} \in (\mathtt{A}^{B(e,R)})^n$ with this property is about $\exp\left[n\,\mathrm{H}\left(\mu^R\right)\right]$. So

$$\inf_{\varepsilon>0} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}_{\sigma\sim\zeta_n} |\Omega(\sigma, \mu, R, \varepsilon)| \leq \mathrm{H}\left(\mu^R\right).$$

Idea of proof

continued – duplication of information

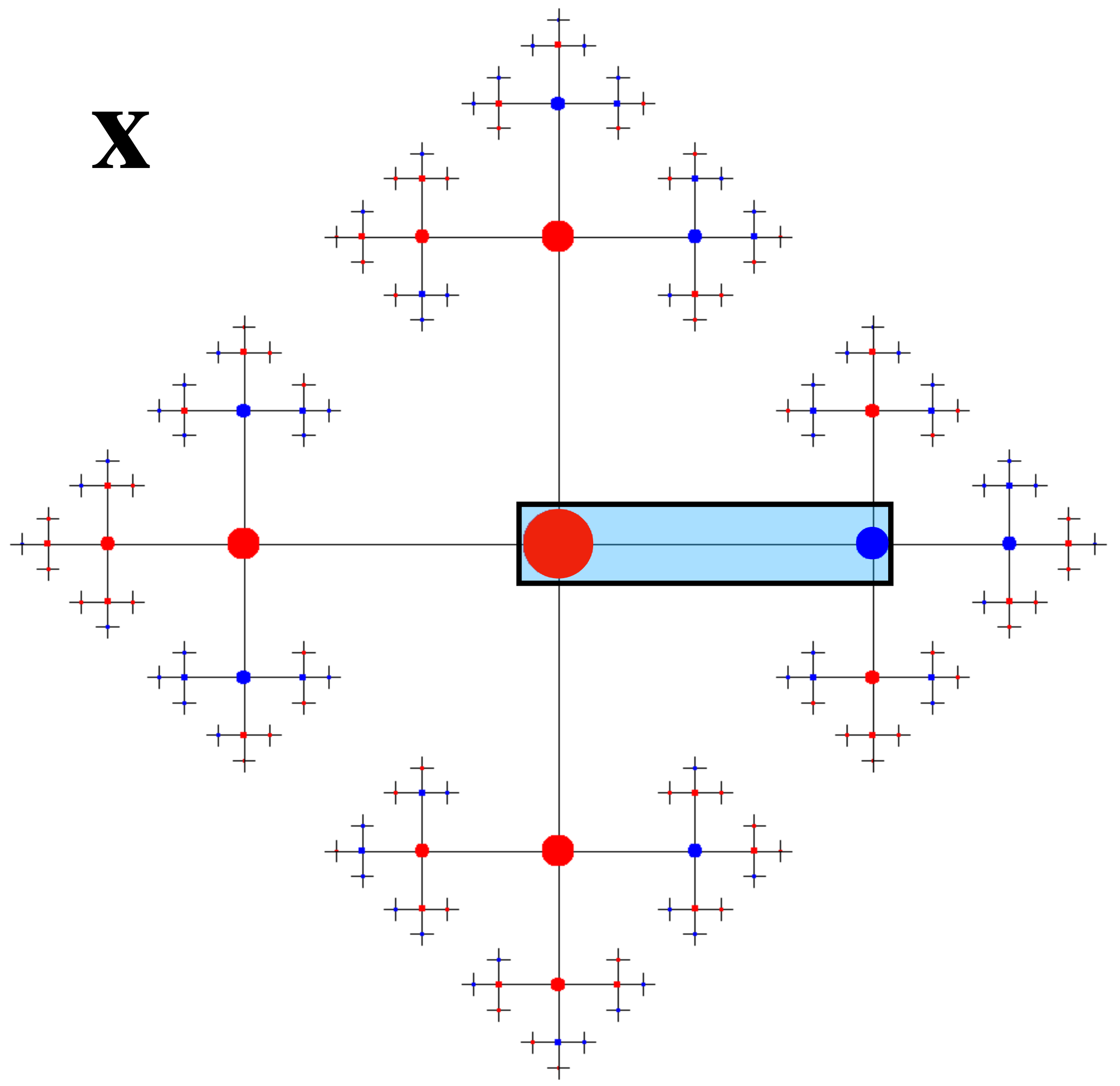


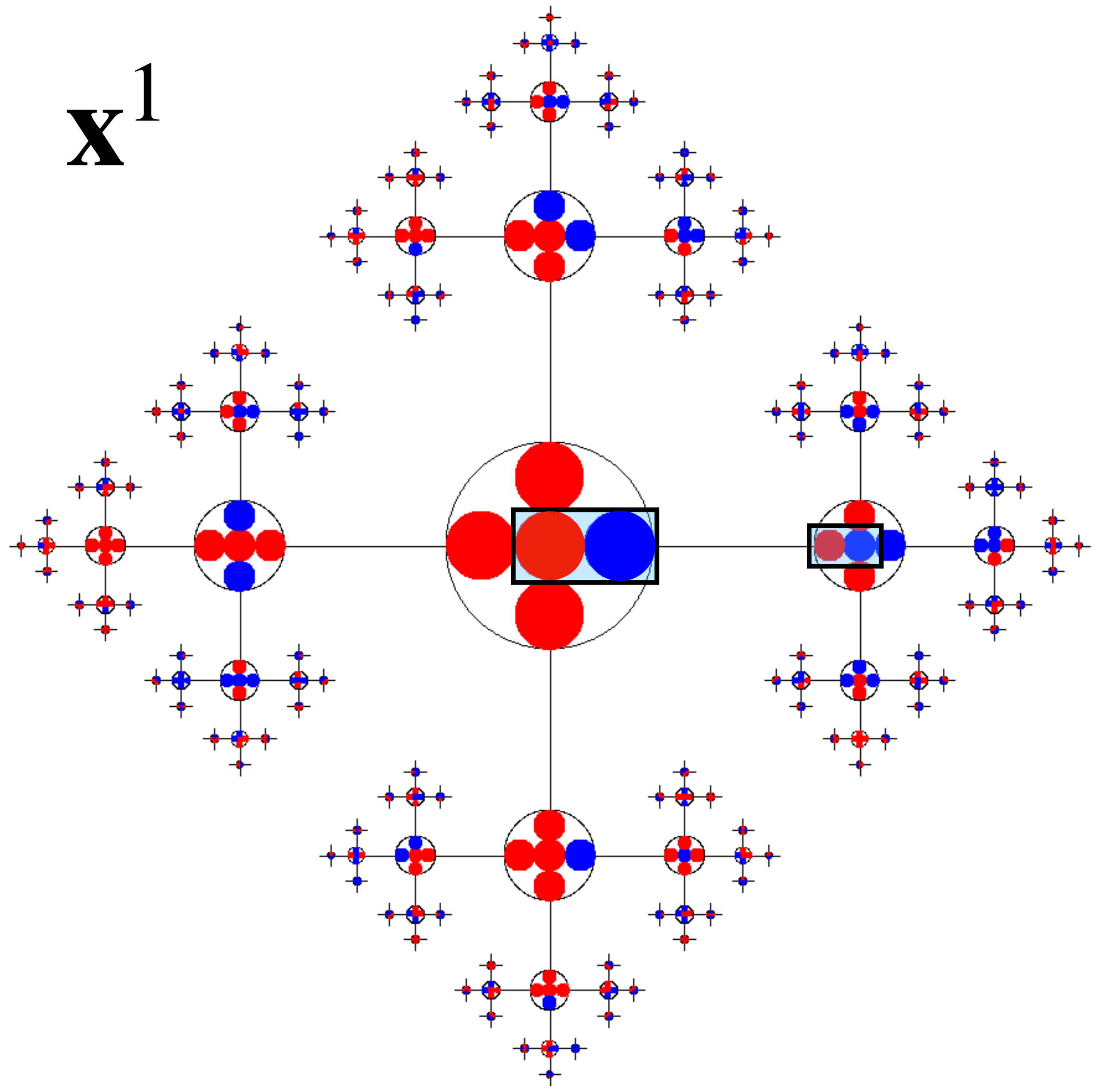


Idea of proof

continued – duplication of information

What if we instead require



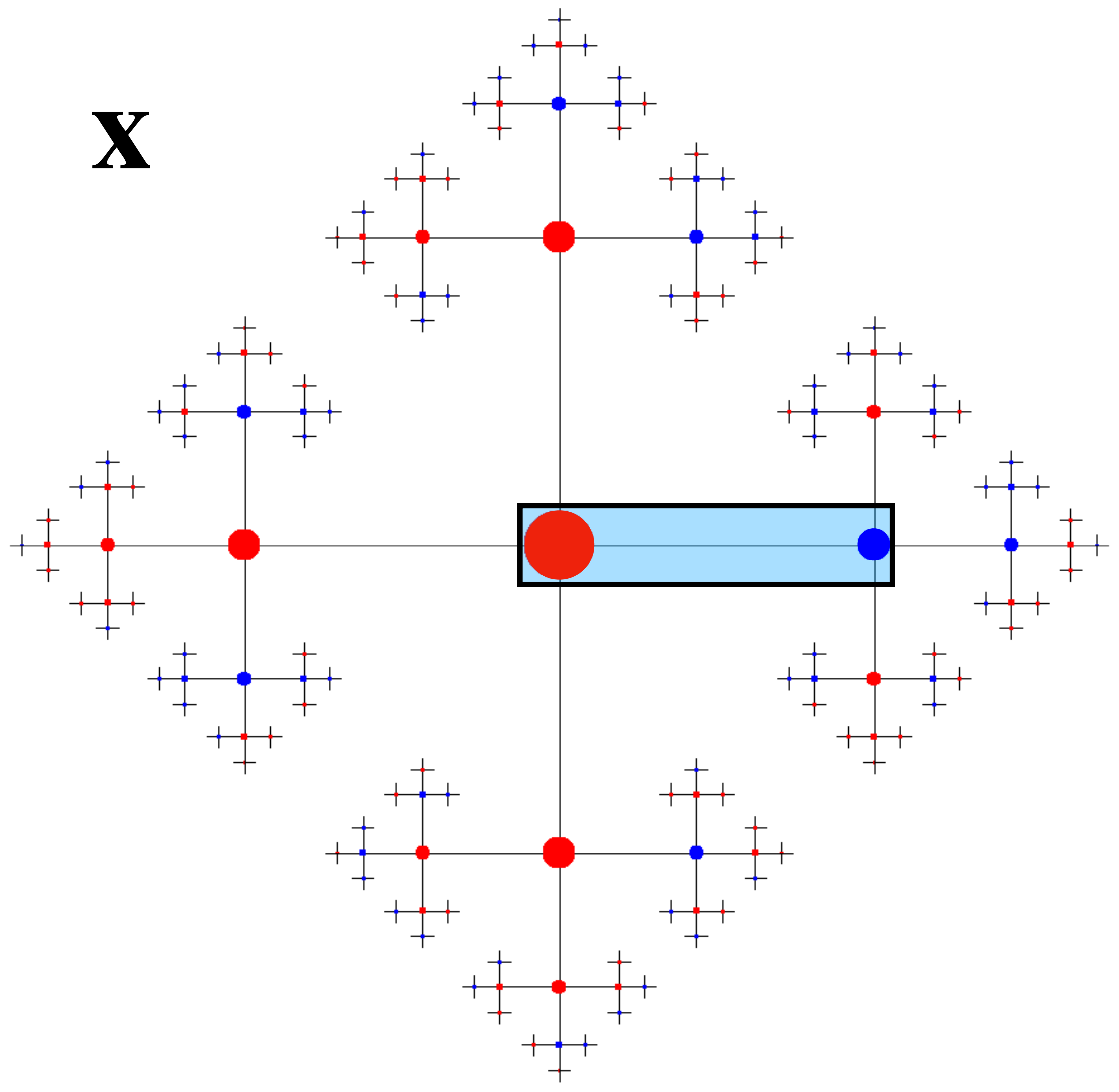


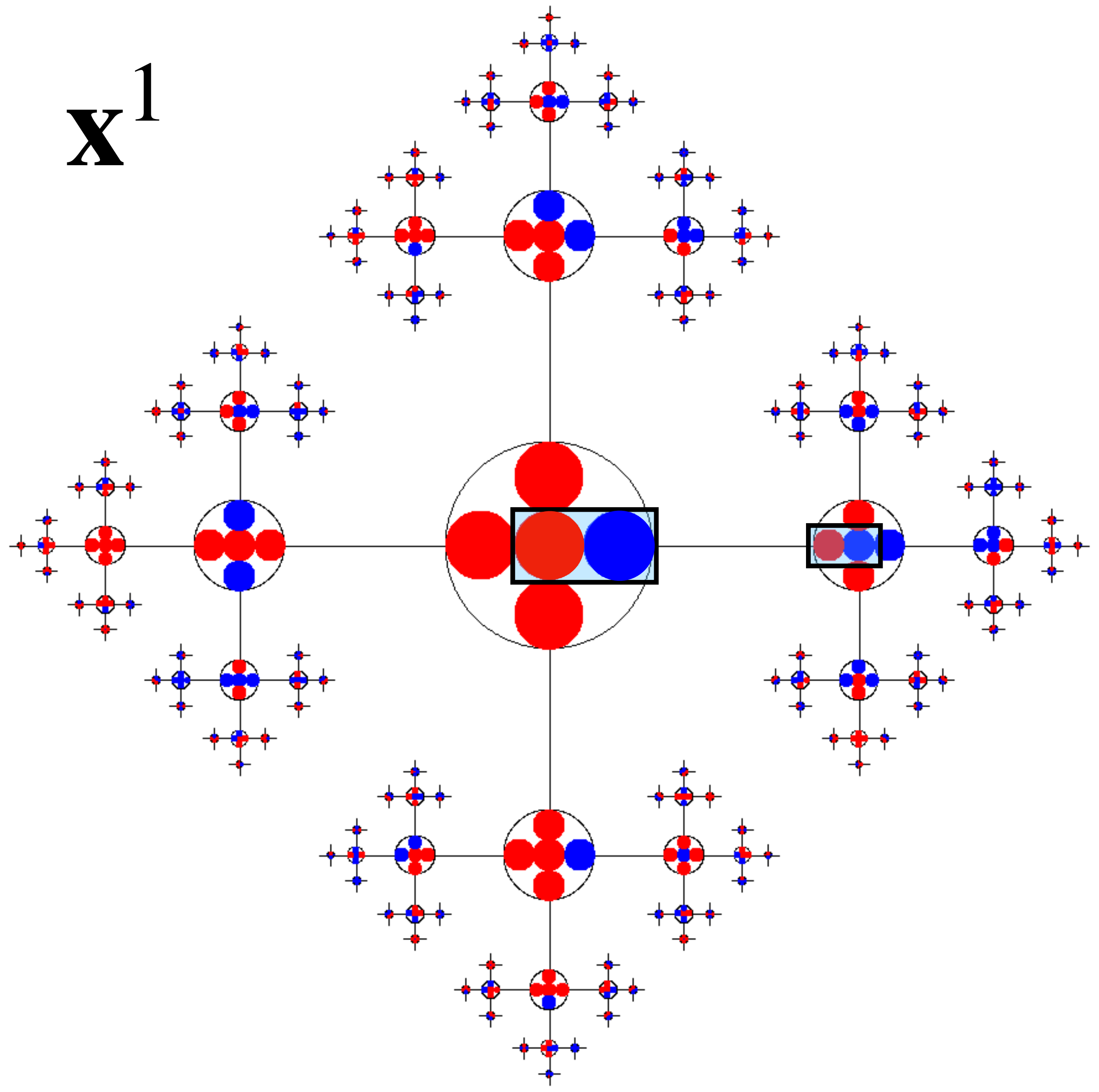
Idea of proof

continued – duplication of information

What if we instead require

$$\frac{1}{n}\#\left\{ j \in [n] : \begin{array}{c} \mathbf{X}(j) = \mathbf{a} \\ \mathbf{X}(\sigma_i j) = \mathbf{a}' \end{array} \right\} \approx \mu \left\{ \begin{array}{c} \mathbf{z}^R(e) = \mathbf{a} \\ \mathbf{z}^R(s_i) = \mathbf{a}' \end{array} \right\}$$





Idea of proof

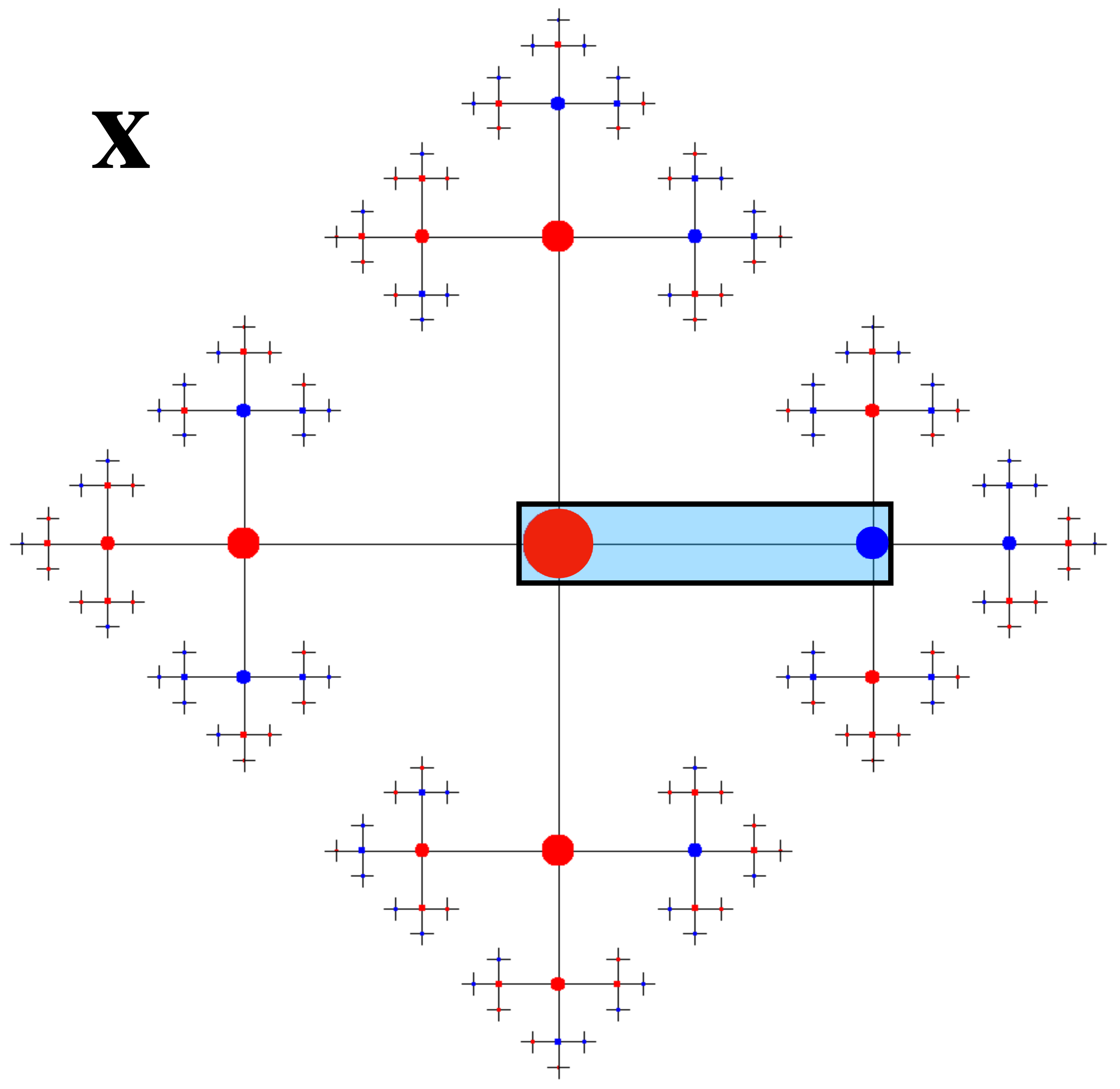
continued – duplication of information

What if we instead require

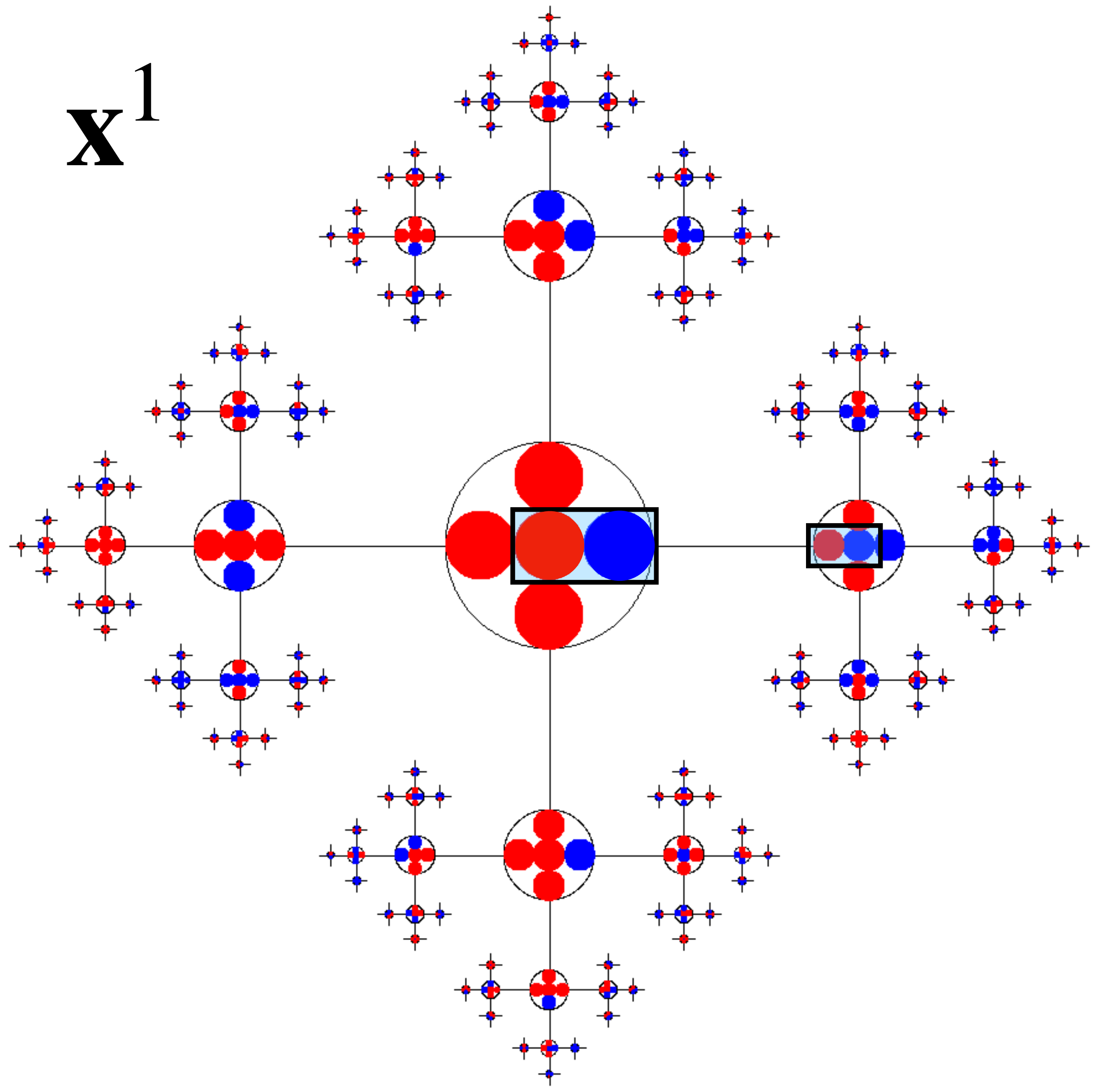
$$\frac{1}{n}\#\left\{ j \in [n] : \begin{array}{c} \mathbf{X}(j) = \mathbf{a} \\ \mathbf{X}(\sigma_i j) = \mathbf{a}' \end{array} \right\} \approx \mu \left\{ \begin{array}{c} \mathbf{z}^R(e) = \mathbf{a} \\ \mathbf{z}^R(s_i) = \mathbf{a}' \end{array} \right\}$$

for all $i \in [r]$?

$\mathbf{x}$



$\mathbf{x}^1$



Idea of proof

continued – duplication of information

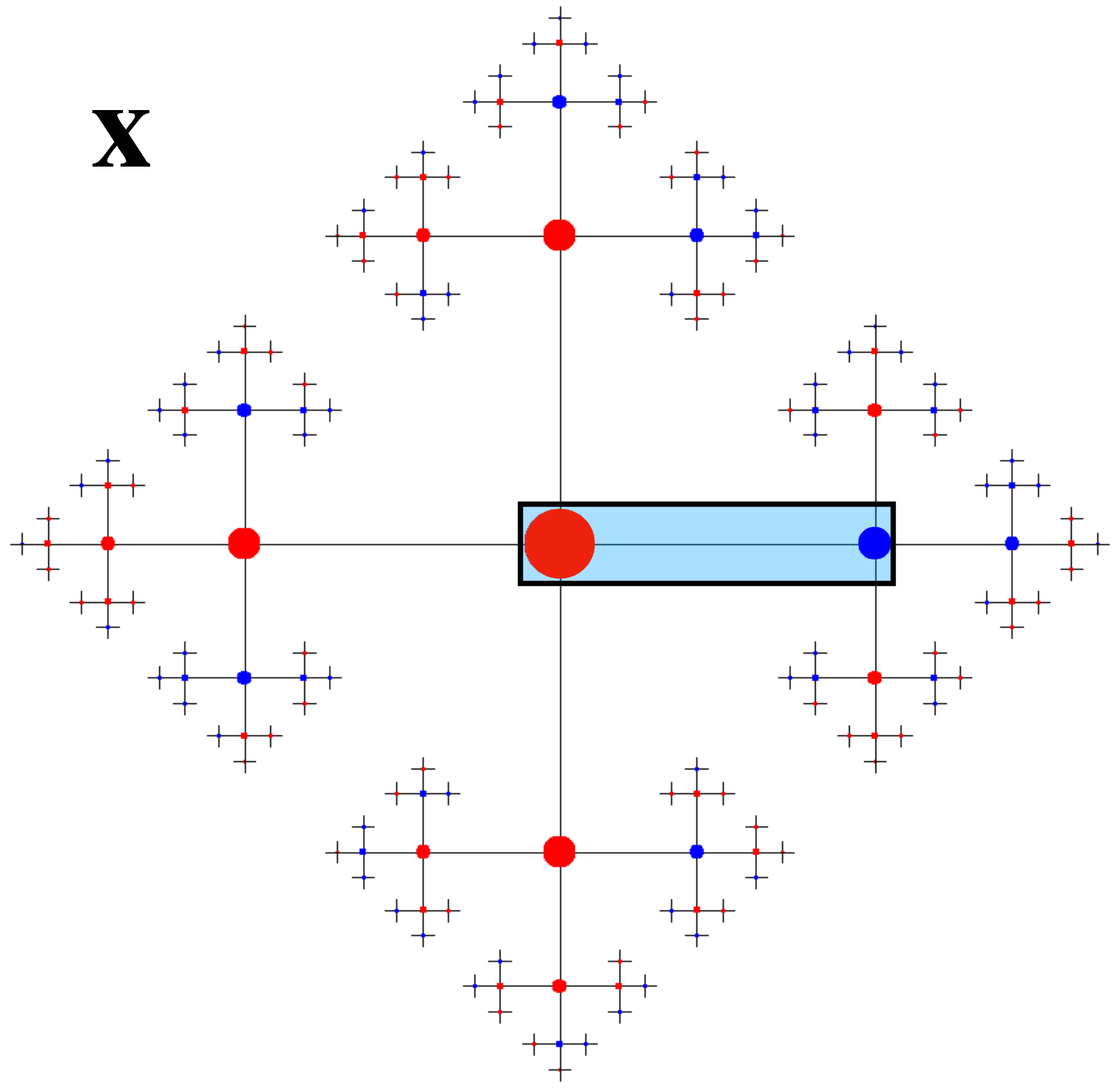
What if we instead require

$$\frac{1}{n}\#\left\{ j \in [n] : \begin{matrix} \mathbf{X}(j) = \mathbf{a} \\ \mathbf{X}(\sigma_i j) = \mathbf{a}' \end{matrix} \right\} \approx \mu \left\{ \begin{matrix} \mathbf{z}^R(e) = \mathbf{a} \\ \mathbf{z}^R(s_i) = \mathbf{a}' \end{matrix} \right\}$$

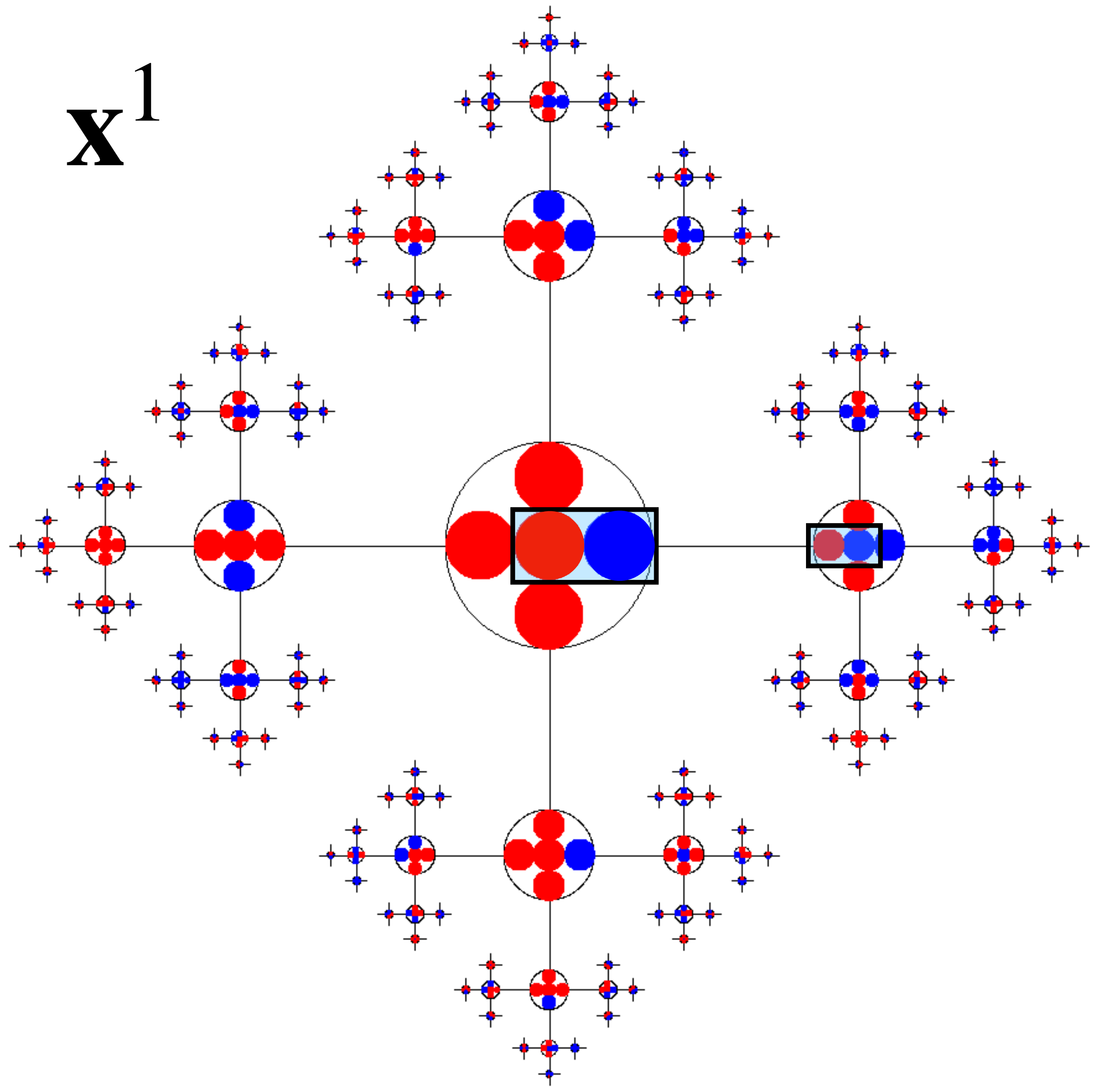
for all $i \in [r]$?

If we take $\mathbf{X} = \mathbf{x}^R$, as a condition on $\mathbf{x}$ this is between $P_{\mathbf{x}}^{\sigma,R} \approx \mu^R$ and $P_{\mathbf{x}}^{\sigma,R+1} \approx \mu^{R+1}$.

$\mathbf{x}$



$\mathbf{x}^1$



Idea of proof

continued – duplication of information

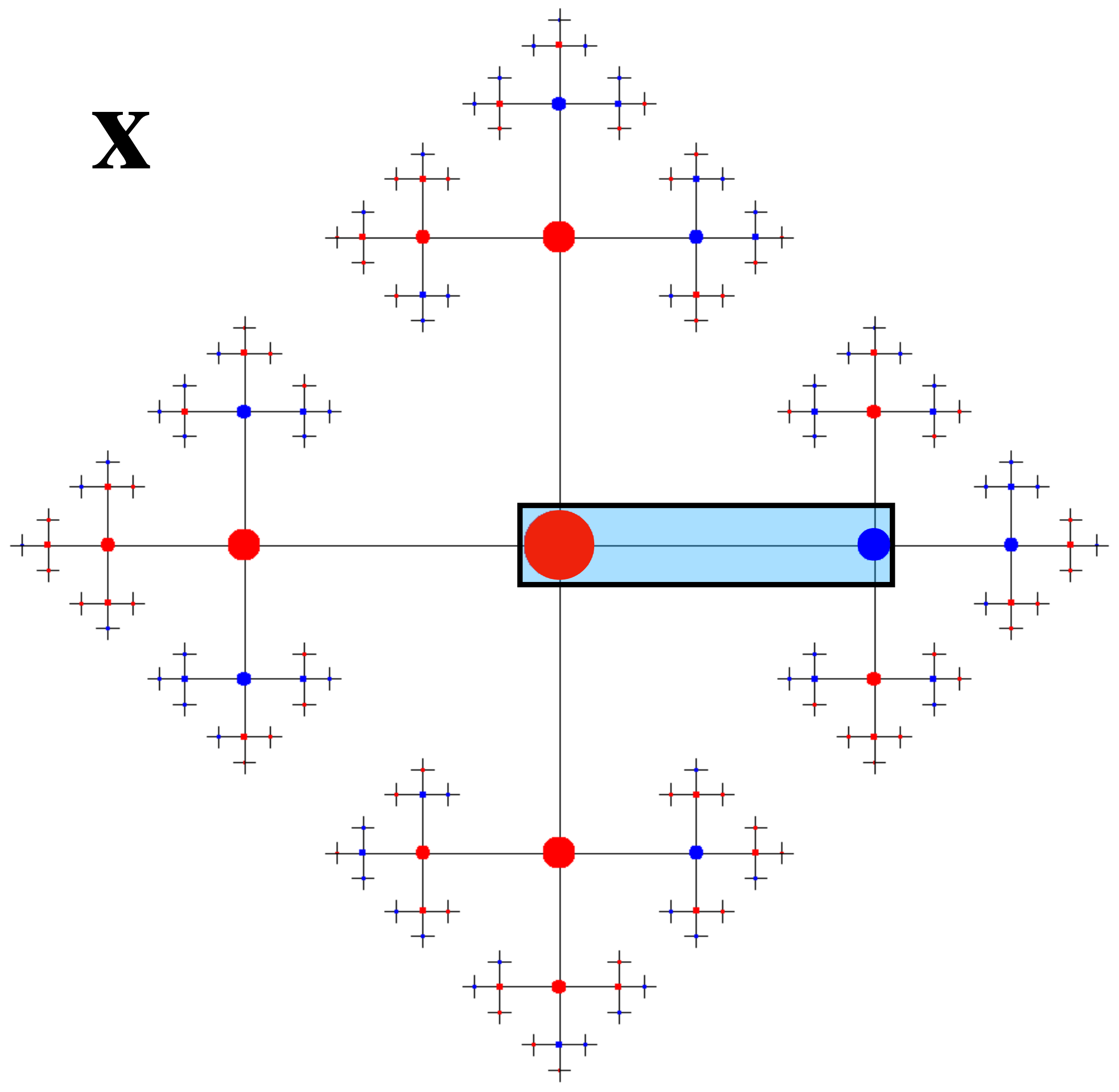
What if we instead require

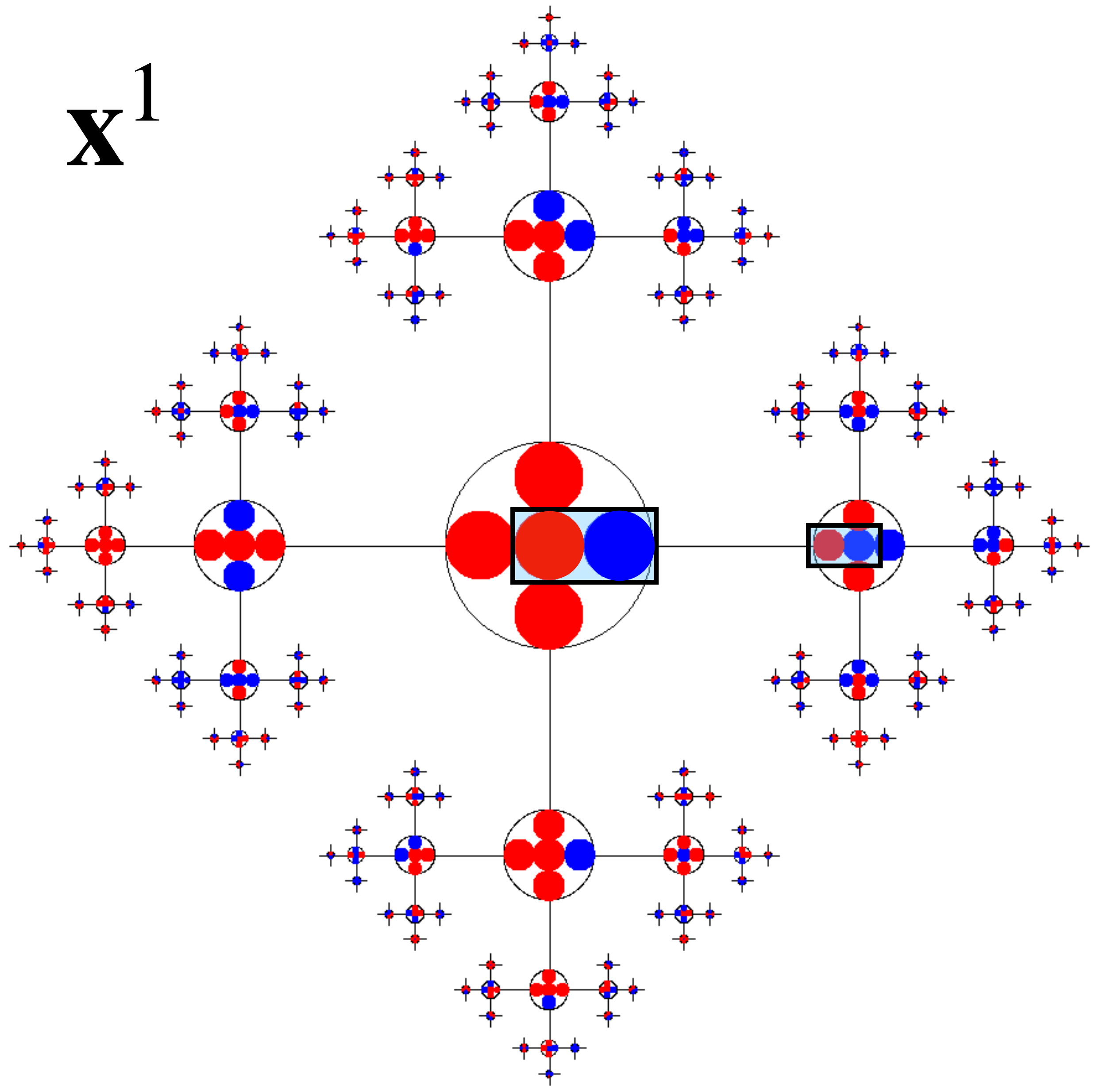
$$\frac{1}{n}\#\left\{ j\in[n] : \begin{matrix}\mathbf{X}(j)=\mathbf{a} \\ \mathbf{X}(\sigma_i j)=\mathbf{a}'\end{matrix} \right\} \approx \mu\left\{ \begin{matrix}\mathbf{z}^R(e)=\mathbf{a} \\ \mathbf{z}^R(s_i)=\mathbf{a}'\end{matrix} \right\}$$

for all $i\in[r]$?

If we take $\mathbf{X}=\mathbf{x}^R$, as a condition on $\mathbf{x}$ this is between $P^{\sigma,R}_{\mathbf{x}}\approx\mu^R$ and $P^{\sigma,R+1}_{\mathbf{x}}\approx\mu^{R+1}$.

So it doesn't affect our notion of "good model" – we're still looking at "local statistics"





Weights

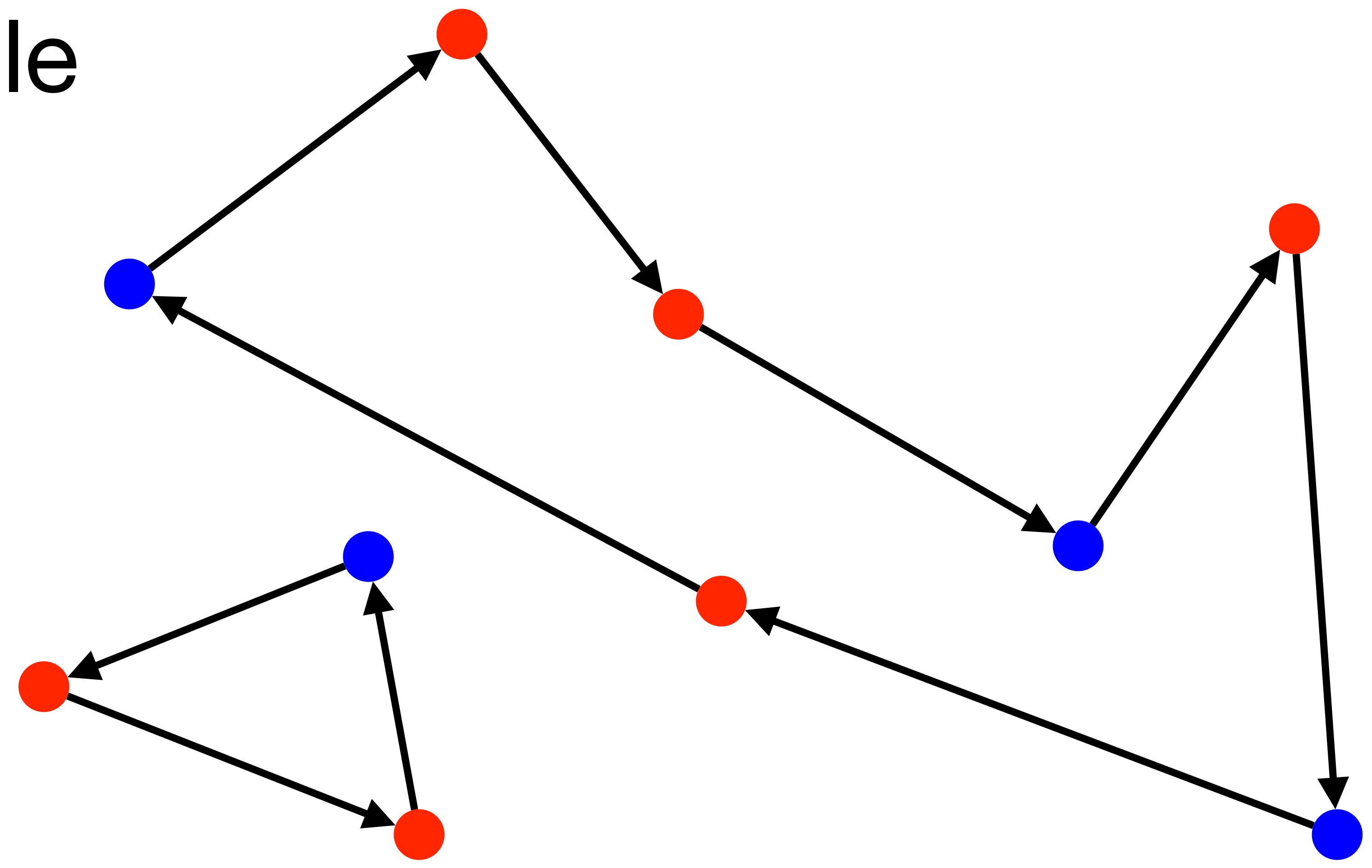
recording "one-step statistics"

For $\tau \in \mathrm{Sym}(n)$, $\mathbf{x} \in \mathtt{A}^n$ let $W_{\mathbf{x},\tau} \in \mathrm{Prob}(\mathtt{A} \times \mathtt{A})$ be given by

$$W_{\mathbf{x},\tau}(\mathsf{a}, \mathsf{a}') = \frac{1}{n} \# \left\{ j \in [n] \ : \ \begin{array}{c} \mathbf{x}(j) = \mathsf{a} \\ \mathbf{x}(\tau j) = \mathsf{a}' \end{array} \right\}.$$

Note both marginals of $W_{\mathbf{x},\tau}$ are equal to $P^0_{\mathbf{x}}$.

Example



$\mathtt{A} = \{ {\color{red}\bullet}, {\color{blue}\bullet} \}$

$p({\color{red}\bullet}) = \frac{3}{5}$

$p({\color{blue}\bullet}) = \frac{2}{5}$

$W({\color{red}\bullet}, {\color{red}\bullet}) = \frac{1}{5}$

$W({\color{red}\bullet}, {\color{blue}\bullet}) = \frac{2}{5}$

$W({\color{blue}\bullet}, {\color{blue}\bullet}) = 0$

$W({\color{blue}\bullet}, {\color{red}\bullet}) = \frac{2}{5}$

Lemma 2

mutual information and counting

Suppose $p \in \mathrm{Prob}(\mathtt{A})$ and that $\lambda \in \mathrm{Prob}(\mathtt{A} \times \mathtt{A})$ is a coupling of p with itself.
Suppose $\mathbf{x} \in \mathtt{A}^n$ has $P^0_{\mathbf{x}} \approx p$.

Then the **proportion** of $\tau \in \mathrm{Sym}(n)$ with $W_{\mathbf{x},\tau} \approx \lambda$ is about

$$\exp\big[-n\,\mathrm{I}(\lambda)\big].$$

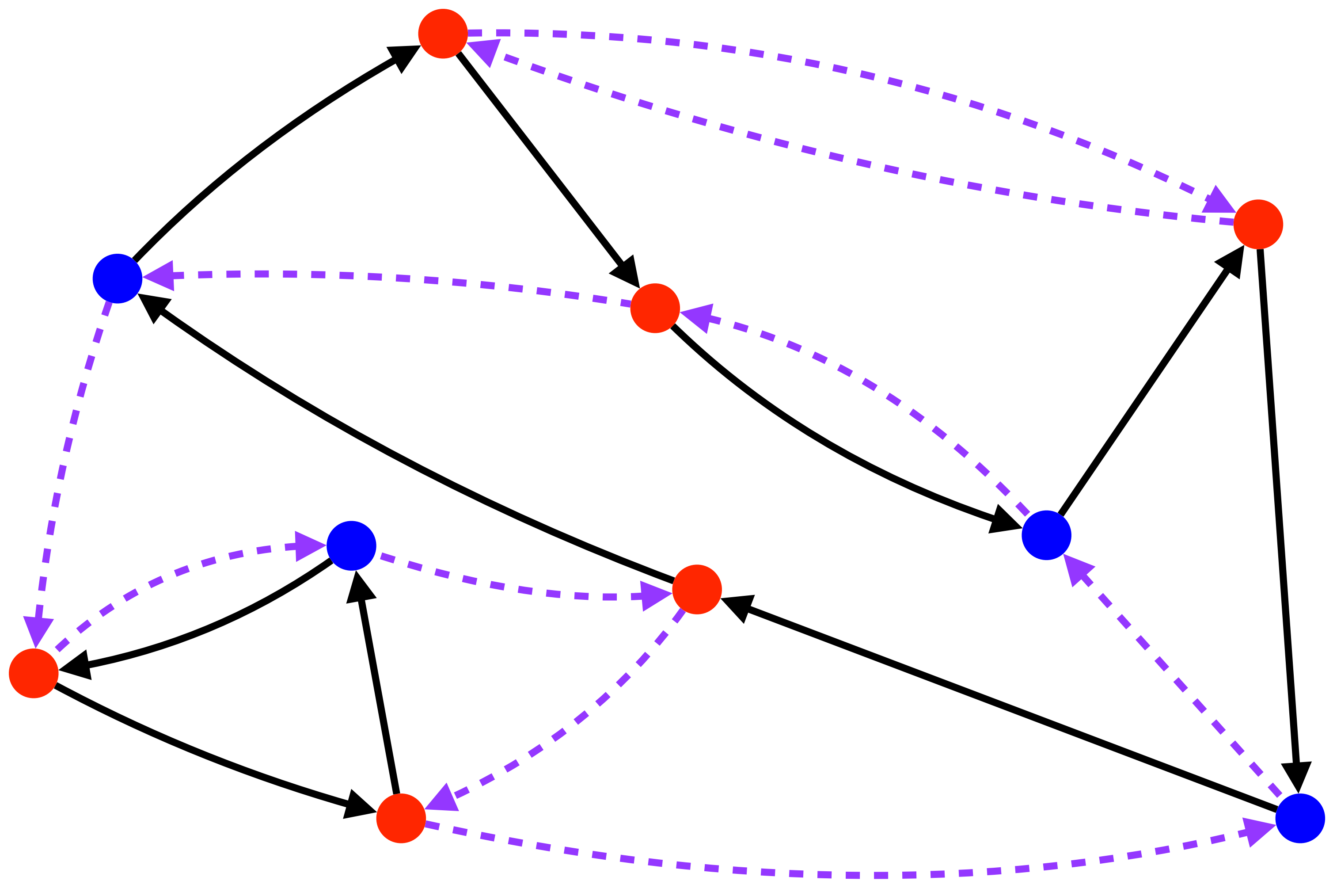
More weights

For $\sigma = (\sigma_1, \ldots, \sigma_r)$, $\mathbf{x} \in \mathtt{A}^n$ let $W_{\mathbf{x},\sigma} \in \mathrm{Prob}(\mathtt{A} \times \mathtt{A})^r$ be given by

$$W_{\mathbf{x},\sigma}(\mathsf{a}, \mathsf{a}'; i) = \frac{1}{n} \# \left\{ j \in [n] \; : \; \begin{array}{c} \mathbf{x}(j) = \mathsf{a} \\ \mathbf{x}(\sigma_i\, j) = \mathsf{a}' \end{array} \right\}.$$

For each $i \in [r]$, both marginals of $W_{\mathbf{x},\sigma}(\,\cdot\,,\,\cdot\,; i)$ are equal to $P^0_{\mathbf{x}}$.

Example



$\mathtt{A} = \{ \textcolor{red}{\bullet}, \textcolor{blue}{\bullet} \}$

$p(\textcolor{red}{\bullet}) = \frac{3}{5}$

$p(\textcolor{blue}{\bullet}) = \frac{2}{5}$

$W(\textcolor{red}{\bullet}, \textcolor{red}{\bullet}; \uparrow) = \frac{1}{5}$ $\quad W(\textcolor{red}{\bullet}, \textcolor{red}{\bullet}; \textcolor{purple}{\uparrow}) = \frac{3}{10}$

$W(\textcolor{red}{\bullet}, \textcolor{blue}{\bullet}; \uparrow) = \frac{2}{5}$ $\quad W(\textcolor{red}{\bullet}, \textcolor{blue}{\bullet}; \textcolor{purple}{\uparrow}) = \frac{3}{10}$

$W(\textcolor{blue}{\bullet}, \textcolor{blue}{\bullet}; \uparrow) = 0$ $\quad W(\textcolor{blue}{\bullet}, \textcolor{blue}{\bullet}; \textcolor{purple}{\uparrow}) = \frac{1}{10}$

$W(\textcolor{blue}{\bullet}, \textcolor{red}{\bullet}; \uparrow) = \frac{2}{5}$ $\quad W(\textcolor{blue}{\bullet}, \textcolor{red}{\bullet}; \textcolor{purple}{\uparrow}) = \frac{3}{10}$

More weights

- For $\mathbf{z} \sim \mu \in \mathrm{Prob}(\mathtt{A}^{\mathbb{F}_r})$ we use a slightly different notation:
- let $W_{\mathbf{z}} \in \mathrm{Prob}(\mathtt{A} \times \mathtt{A})^r$ be given by

$$W_{\mathbf{z}}(\mathsf{a}, \mathsf{a}'; i) = \mu \begin{Bmatrix} \mathbf{z}(e) = \mathsf{a} \\ \mathbf{z}(s_i) = \mathsf{a}' \end{Bmatrix}.$$

- So:
  - $W_{\mathbf{z}}(\,\cdot\,,\,\cdot\,;i) = \mathrm{Law}\big(\mathbf{z}(e), \mathbf{z}(s_i)\big)$
  - For each i, both marginals of $W_{\mathbf{z}}(\,\cdot\,,\,\cdot\,;i)$ are μ^0.

Similarly, write

$$W_{\mathbf{z}^R}(\,\cdot\,,\,\cdot\,;i) = \mathrm{Law}\big(\mathbf{z}^R(e), \mathbf{z}^R(s_i)\big)$$

Corollary
of Lemma 2

- Let $\mathbf{X} \in (\mathbb{A}^{B(e,R)})^n$ be such that $P^0_{\mathbf{X}} \approx \mu^R$.
- Again write $\mathbf{z} \sim \mu$.
- Then the proportion of $\tau = (\tau_1, \ldots, \tau_r) \in \mathrm{Sym}(n)^r$ with $W_{\mathbf{X},\tau} \approx W_{\mathbf{z}^R}$ is about
$$\prod_{i \in [r]} \exp\big[-n\, \mathrm{I}\big(\mathbf{z}^R(e);\, \mathbf{z}^R(s_i)\big)\big].$$
- *Proof*: let $\lambda_i = W_{\mathbf{z}^R}(\,\cdot\,,\,\cdot\,;i)$ and apply Lemma 2. The key is that $\tau_1, \ldots, \tau_r$ are independent when τ is chosen uniformly

Idea of proof

upper bound sketch

$$\mathbb{E}_\sigma\left[\#\left\{\mathbf{x} \in \mathtt{A}^n : W_{\mathbf{x}^R,\sigma} \approx W_{\mathbf{z}^R}\right\}\right] \le \mathbb{E}_\sigma\left[\#\left\{\mathbf{X} \in (\mathtt{A}^{B(e,R)})^n : W_{\mathbf{X},\sigma} \approx W_{\mathbf{z}^R}\right\}\right] \qquad \mathbf{x} \hookrightarrow \mathbf{x}^R$$

$$= \sum_{\mathbf{X}:\ P^0_{\mathbf{X}} \approx \mu^R} \mathbb{P}_\sigma\left[W_{\mathbf{X},\sigma} \approx W_{\mathbf{z}^R}\right] \qquad \text{linearity of expectation}$$

$$\approx \sum_{\mathbf{X}:\ P^0_{\mathbf{X}} \approx \mu^R} \prod_{i \in [r]} \exp\left[-n\,\mathrm{I}\left(\mathbf{z}^R(e); \mathbf{z}^R(s_i)\right)\right] \qquad \text{Corollary}$$

$$\approx \exp\left(n\,\mathrm{H}(\mu^R)\right) \times \prod_{i \in [r]} \exp\left[-n\,\mathrm{I}\left(\mathbf{z}^R(e); \mathbf{z}^R(s_i)\right)\right] \qquad \text{Lemma 1}$$

$\Rightarrow f(\mu) \le \mathrm{H}\left(\mathbf{z}^R(e)\right) - \sum_{i \in [r]} \mathrm{I}\left(\mathbf{z}^R(e);\ \mathbf{z}^R(s_i)\right)$. Taking the inf over R gives upper bound.

3. Relative f-invariant

Based on:
C. Shriver. "The Relative f-Invariant and Non-Uniform Random Sofic Approximations." Mar. 2, 2020. arXiv: 2003.00663 [math].

Conditional entropy

- Given two coupled random variables $\mathbf{x}, \mathbf{y}$, the **conditional Shannon entropy** is

$$\mathrm{H}(\mathbf{x} \mid \mathbf{y}) = \mathrm{H}(\mathbf{x}, \mathbf{y}) - \mathrm{H}(\mathbf{y}).$$

- We can define a **relative f-invariant**:
  - suppose we have two finite alphabets $\mathtt{A}, \mathtt{B}$ and shift-invariant $\mu_{\mathtt{A}} \in \mathrm{Prob}(\mathtt{A}^{\mathbb{F}_r})$, $\mu_{\mathtt{B}} \in \mathrm{Prob}(\mathtt{B}^{\mathbb{F}_r})$.
  - suppose $\mu \in \mathrm{Prob}\big((\mathtt{A} \times \mathtt{B})^{\mathbb{F}_r}\big)$ is a joining of $\mu_{\mathtt{A}}, \mu_{\mathtt{B}}$. Then

$$f(\mu \mid \mathtt{B}) = f(\mu) - f(\mu_{\mathtt{B}}).$$

Relative f-invariant via good models

Theorem
[S '20]

$$f(\mu \mid \mathtt{B}) = \inf_{R,\varepsilon} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}_{\sigma\sim\mathtt{SBM}_n} \left[\#\left\{ \mathbf{x} \in \mathtt{A}^n : \|P^{\sigma,R}_{(\mathbf{x},\mathbf{y}_n)} - \mu^R\|_{\mathrm{TV}} < \varepsilon \right\} \right]$$

new objects: $\mathtt{SBM}_n$ is a type of **stochastic block model** with a **planted partition** $\mathbf{y}_n$

This encodes the 'already known' information from $\mathtt{B}$

- $\mathtt{SBM}_n$ will be such that $\mathbf{y}_n$ is a good model for $\mu_\mathtt{B}$
- we're counting the expected number of good models for $\mu_\mathtt{A}$ which extend $\mathbf{y}_n$ to μ, a particular joining of $\mu_\mathtt{A}, \mu_\mathtt{B}$

Permutation stochastic block models

Given $\tau = (\tau_1, \ldots, \tau_r) \in \mathrm{Sym}(n)^r$, $\mathbf{y} \in \mathtt{B}^n$, and $k \in \mathbb{N}$ let

$$\mathtt{SBM}(\tau, \mathbf{y}, k) = \mathrm{Unif}\left\{\sigma \in \mathrm{Sym}(n)^r : W_{\mathbf{y}^k,\sigma} = W_{\mathbf{y}^k,\tau} \ \forall i\right\}.$$

planted partition of $[n]$

used to specify local statistics

depends on all of σ or τ

$k = 0 \longrightarrow$ standard SBM

$k > 0 \longrightarrow$ more precise local statistics

Theorem [S '20]

$$f(\mu \mid \mathtt{B}) = \inf_{R,\varepsilon} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}_{\sigma\sim\mathtt{SBM}_n} \left[\#\left\{\mathbf{x} \in \mathtt{A}^n : \|P^{\sigma,R}_{(\mathbf{x},\mathbf{y}_n)} - \mu^R\|_{\mathrm{TV}} < \varepsilon\right\}\right]$$

What parameters work?

- pick $m_n = o(\log\log n)$
- pick $\mathbf{y}_n \in \mathtt{B}^n$ with $P^0_{\mathbf{y}_n} \approx \mu^0_{\mathtt{B}}$ (individual letter frequencies are correct)
- pick τ_n so that $W_{\mathbf{y}_n^{m_n},\tau_n} \approx W_{\mathbf{z}_{\mathtt{B}}^{m_n}}$ (here $\mathbf{z}_{\mathtt{B}} \sim \mu_{\mathtt{B}}$; the radius-$m_n$ weight of $\mathbf{y}_n$ is correct)
- let $\mathtt{SBM}_n = \mathtt{SBM}(\tau_n, \mathbf{y}_n, m_n)$.

For $\approx$'s, precise estimates are needed, but best choices always work.

Good models over an SBM

- What is

$$\inf_{R,\varepsilon} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}_{\sigma\sim\mathtt{SBM}_n} |\Omega(\sigma, \mu_{\mathtt{A}}, R, \varepsilon)| ?$$

- Let $\mathbf{y}_n \in \mathtt{B}^n$ be the planted good model for $\mu_{\mathtt{B}}$ from $\mathtt{SBM}_n$.
  - If $\mathbf{x}$ is any good model for $\mu_{\mathtt{A}}$ then $(\mathbf{x}, \mathbf{y}_n)$ is a good model for *some* joining of $\mu_{\mathtt{A}}, \mu_{\mathtt{B}}$ – maybe not μ.

Good models over an SBM

- Let $\mathscr{J}(\mu_{\mathtt{A}}, \mu_{\mathtt{B}}) \subset \operatorname{Prob}\left((\mathtt{A} \times \mathtt{B})^{\mathbb{F}_r}\right)$ denote the set of joinings of $\mu_{\mathtt{A}}, \mu_{\mathtt{B}}$

Theorem [S '20]

$$\inf_{R,\varepsilon} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}_{\sigma\sim\mathtt{SBM}_n} |\Omega(\sigma, \mu_{\mathtt{A}}, R, \varepsilon)| = \sup_{\lambda\in\mathscr{J}(\mu_{\mathtt{A}},\mu_{\mathtt{B}})} f(\lambda \mid \mathtt{B})$$

- Write LHS as $\mathrm{h}_{\boldsymbol{\Sigma}}(\mu_{\mathtt{A}})$, with $\boldsymbol{\Sigma} = (\mathtt{SBM}_n)_{n=1}^{\infty}$ a random sofic approximation.

Summary and future work

- We have formulas for entropy over two types of random sofic approximations:
  - uniform $\rightarrow f$-invariant
  - stochastic block model $\rightarrow$ optimum over relative f-invariants
- Different entropy values for *nonrandom* sofic approximations?
  - "degenerate" case of no good models is known to occur
- SBM's can avoid degeneracy by ensuring the existence of *some* good models.
  - need to understand optimization better. Some progress for when $\mu_{\mathtt{A}}, \mu_{\mathtt{B}}$ are Gibbs measures for a nearest-neighbor interaction (like Ising).