Size of maximum b-matchings in sparse random graphs

Mathieu Leconte

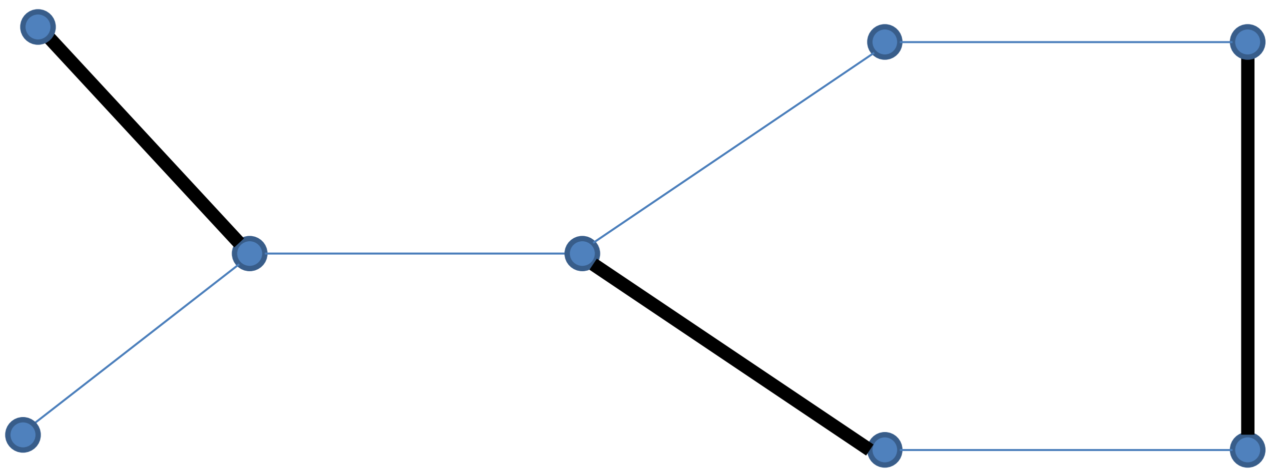
Joint work with Marc Lelarge and Laurent Massoulié

Goal?

- Computing the asymptotic density of a b-matching as the size of graphs tend to infinity
  - Can we compute other quantities in a similar way?
- Sequences of sparse random graphs with a known local structure
  - Graphs that converge (in the local weak sense) towards Galton-Watson trees
    - Ex: Erdos-Renyi, configuration model

Matchings

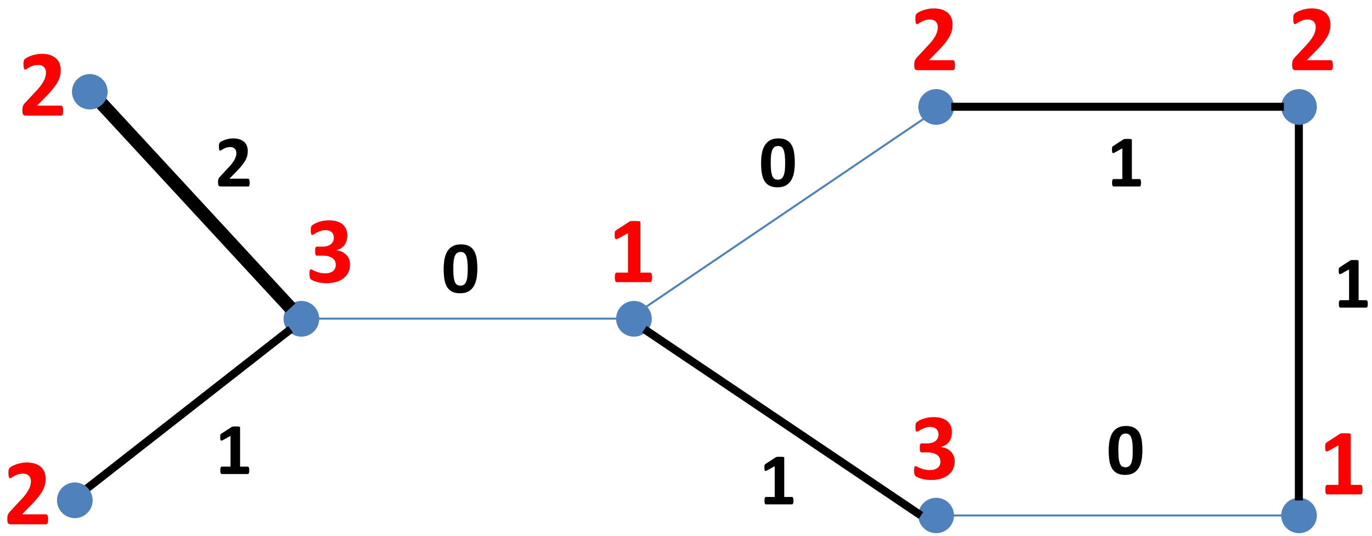
- Graph *G*=(*V*,*E*)
- Matching= subset of edges *E'* such that each vertex is adjacent to at most one element in *E'*



studied by Zdeborova and Mézard (2006)
Bordenave, Lelarge, Salez (2011)

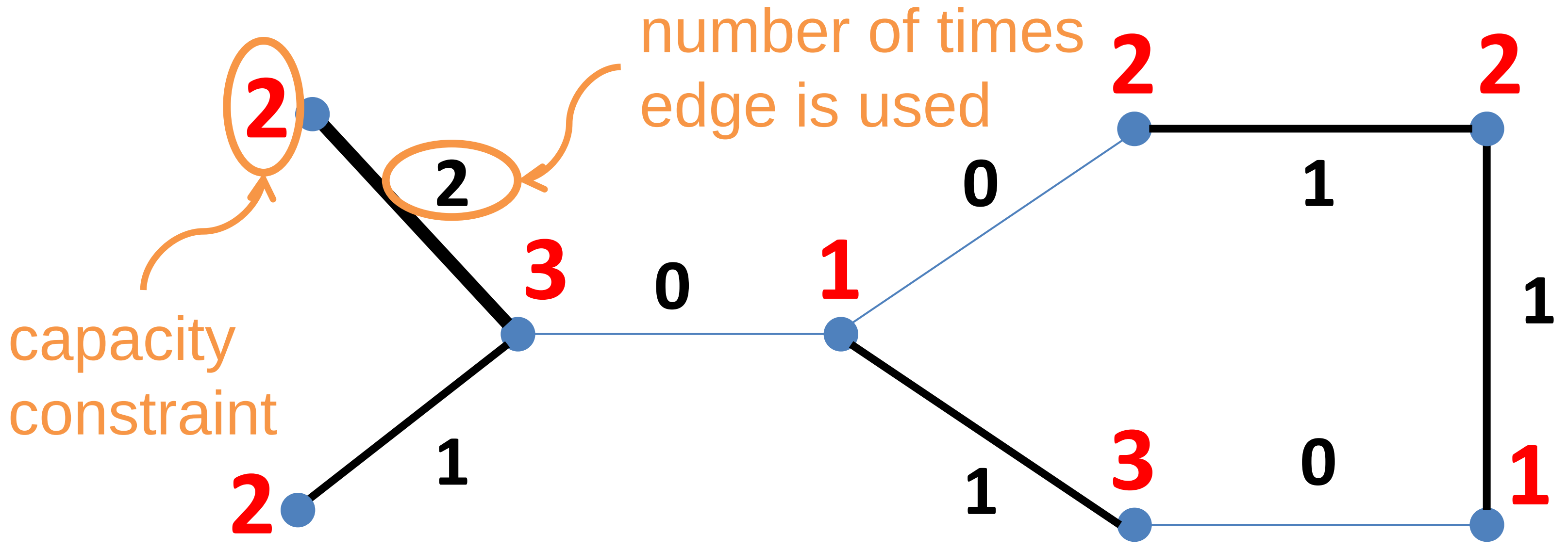
b-matchings

- Capacity constraint b_v at vertex v
- Each edge may be used more than once
- Total usage of edges adjacent to v must not exceed b_v

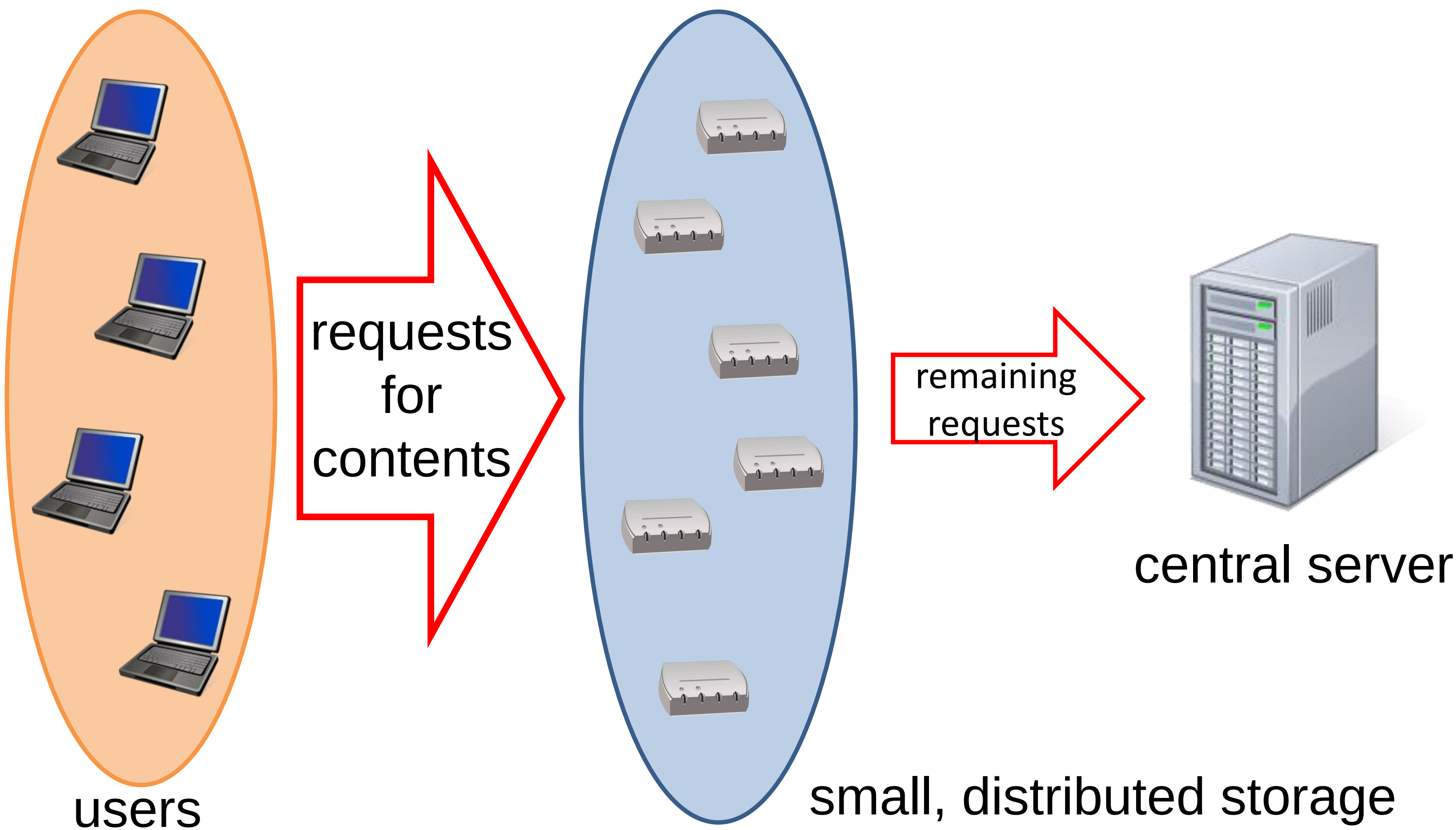


b-matchings

- Capacity constraint b_v at vertex v
- Each edge may be used more than once
- Total usage of edges adjacent to v must not exceed b_v

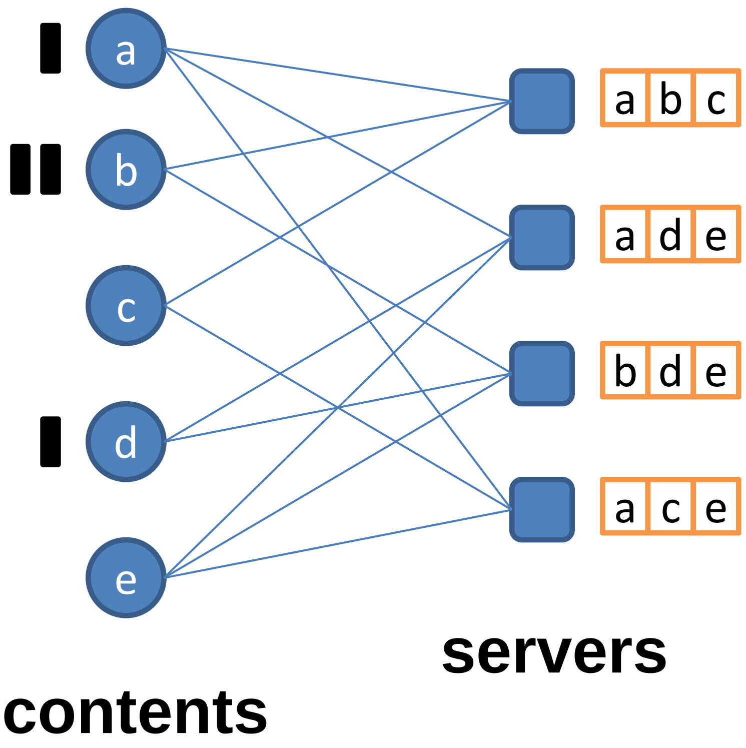


An application: distributed content distribution system



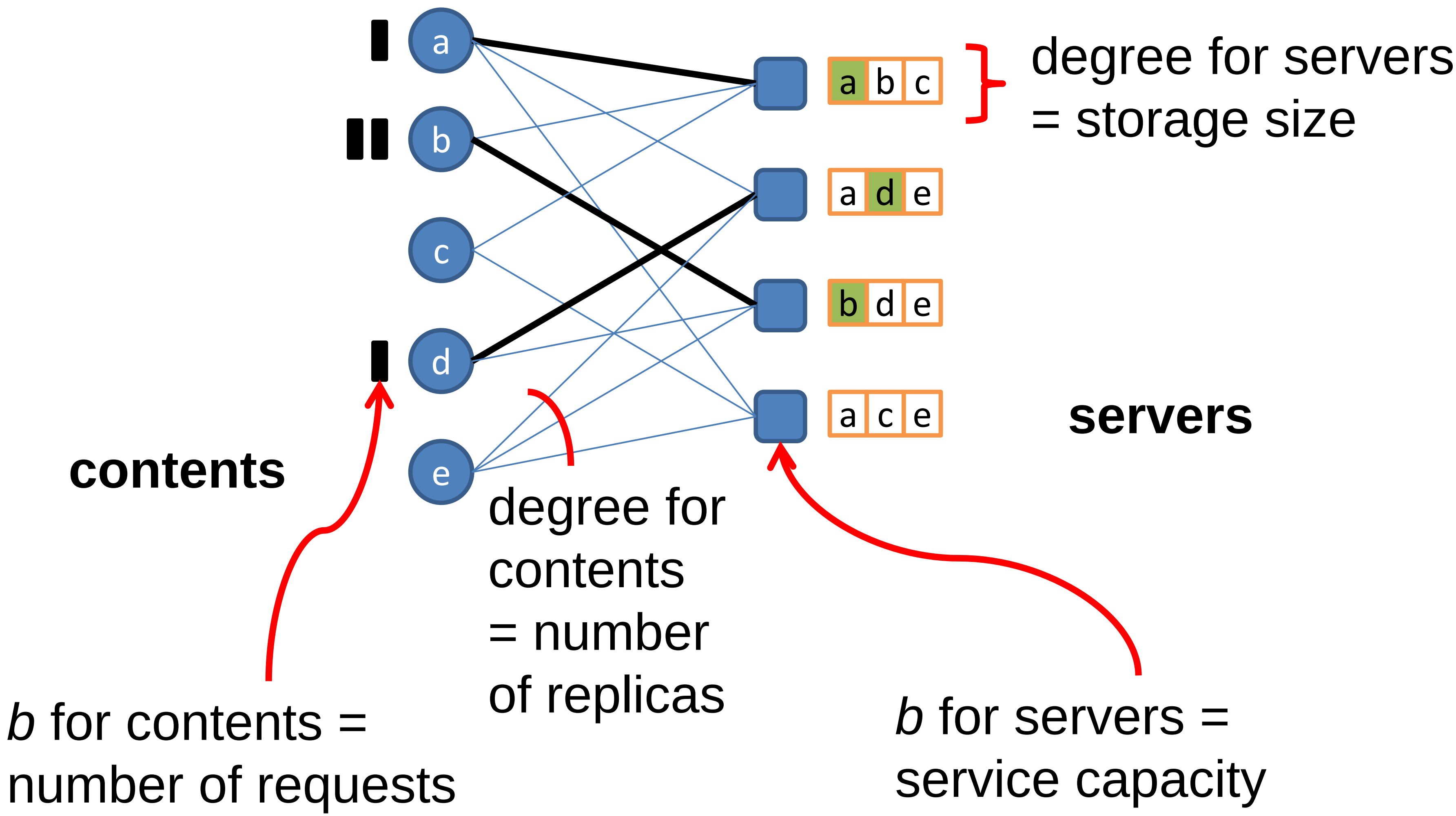
Distributed content distribution system

- *n* contents, *m* servers
- Contents stored by each server determined independently (but not uniformly)
- Number of requests and number of replicas for each content jointly determined
- $n \to \infty, n/m \to \beta$

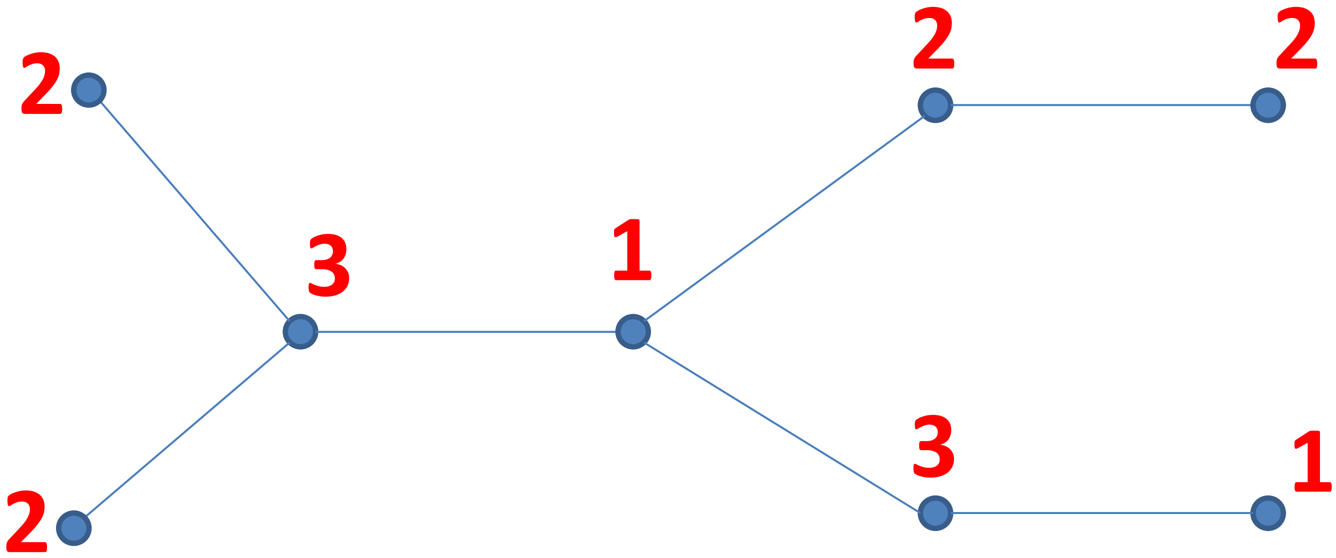


What fraction of the requests can be served?

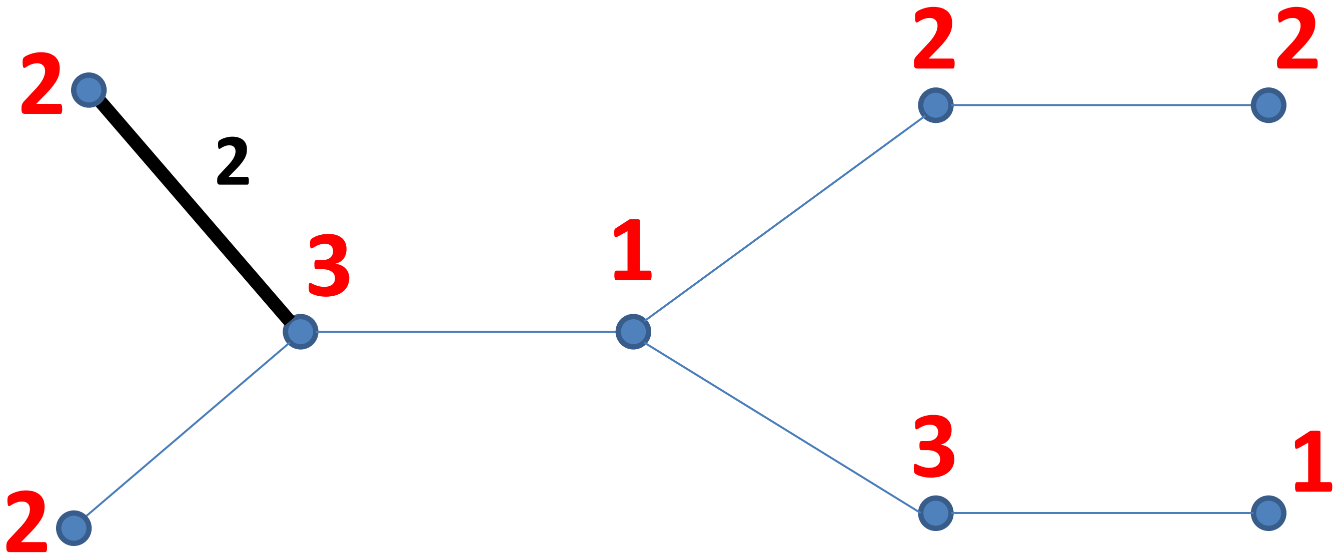
Distributed content distribution system



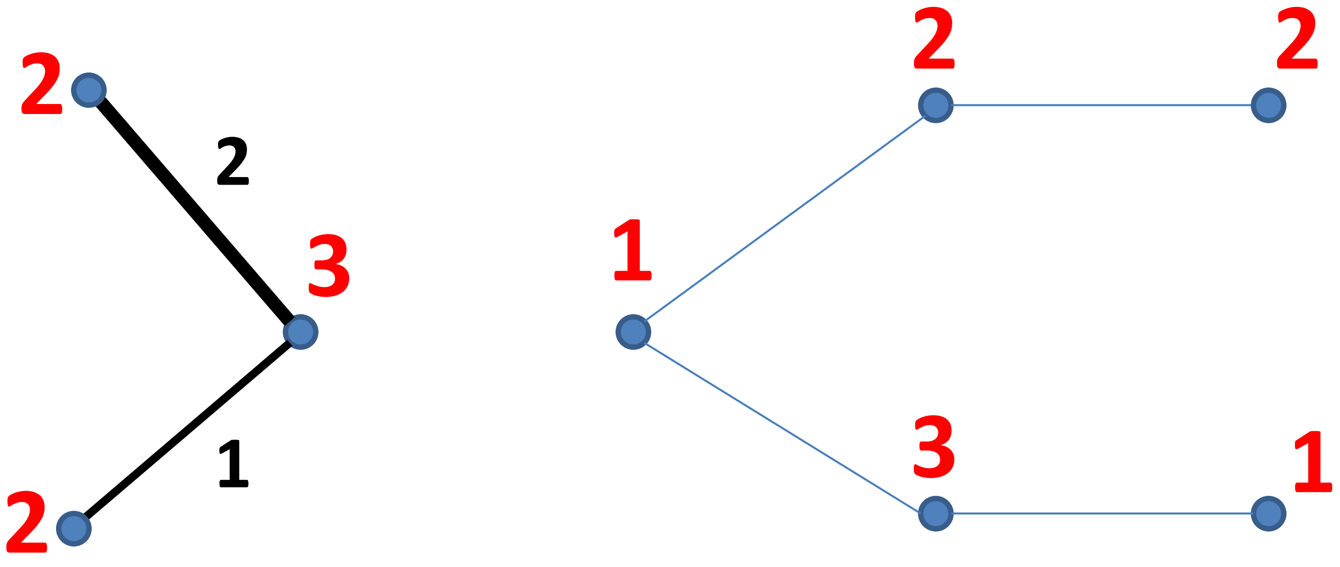
Greedy algorithm on trees for finding a maximum b-matching



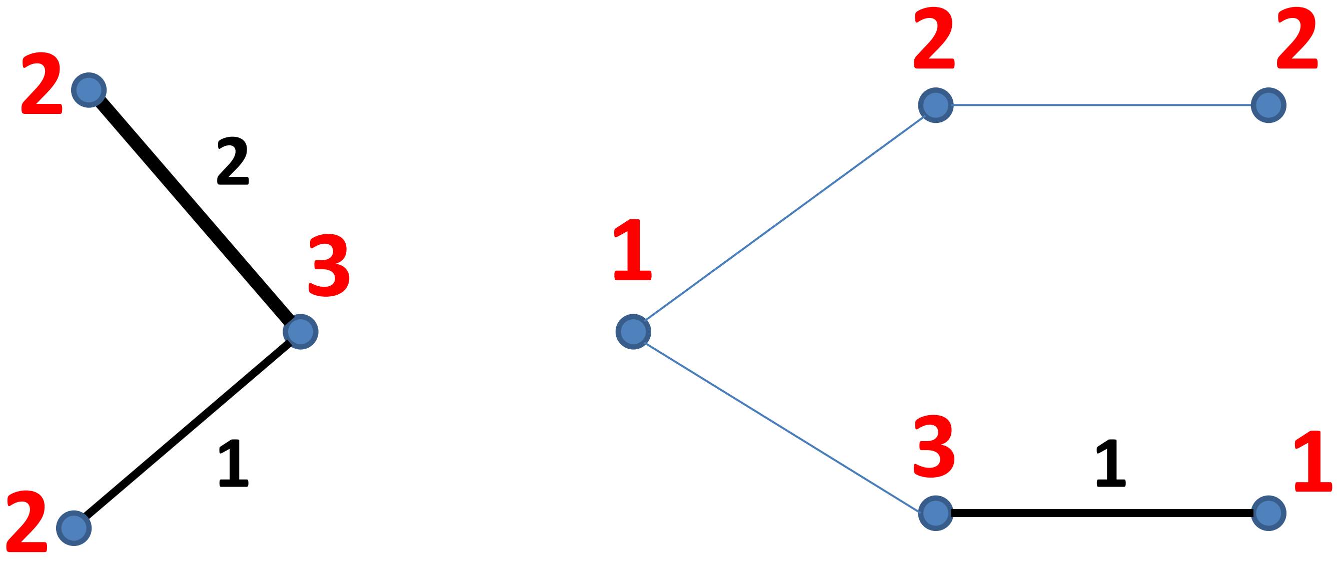
Greedy algorithm on trees



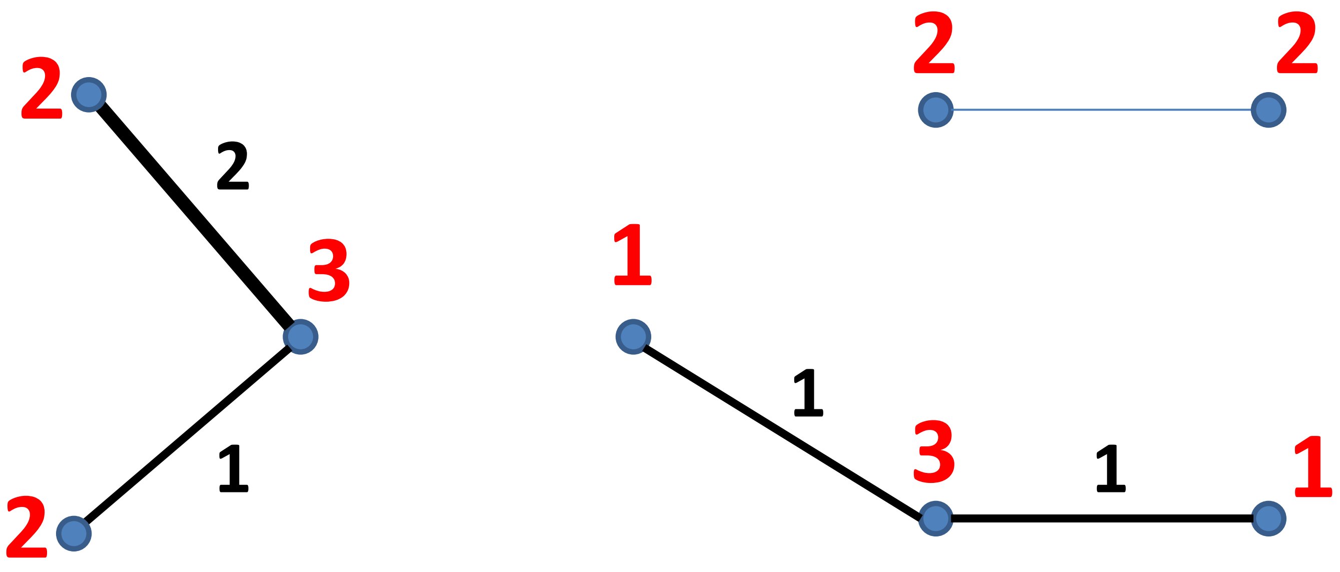
Greedy algorithm on trees



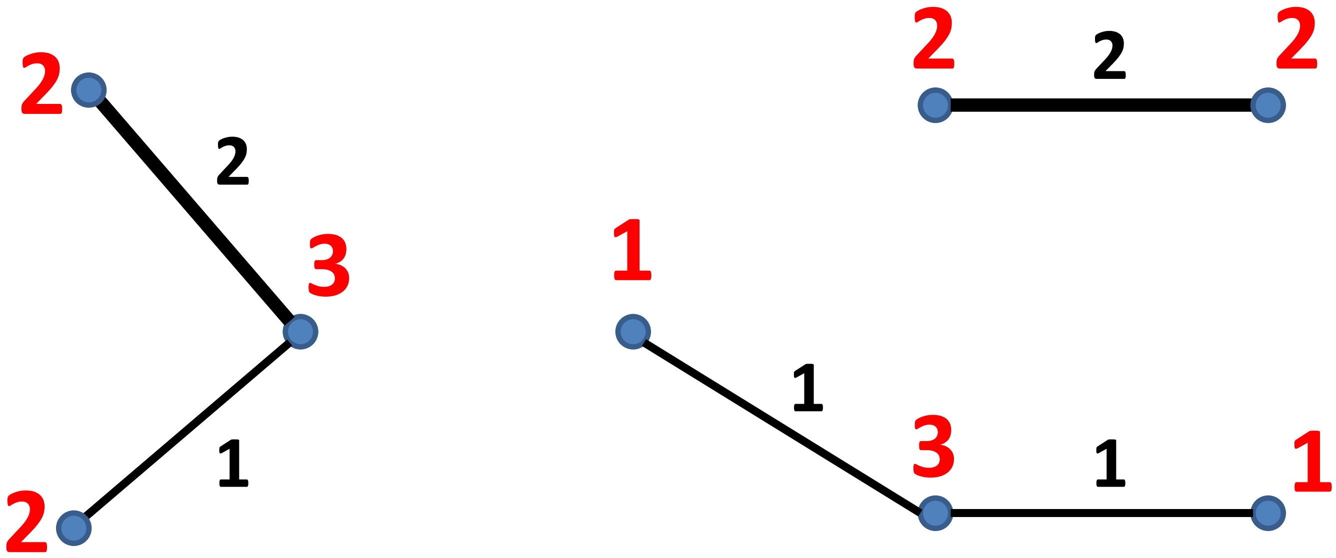
Greedy algorithm on trees



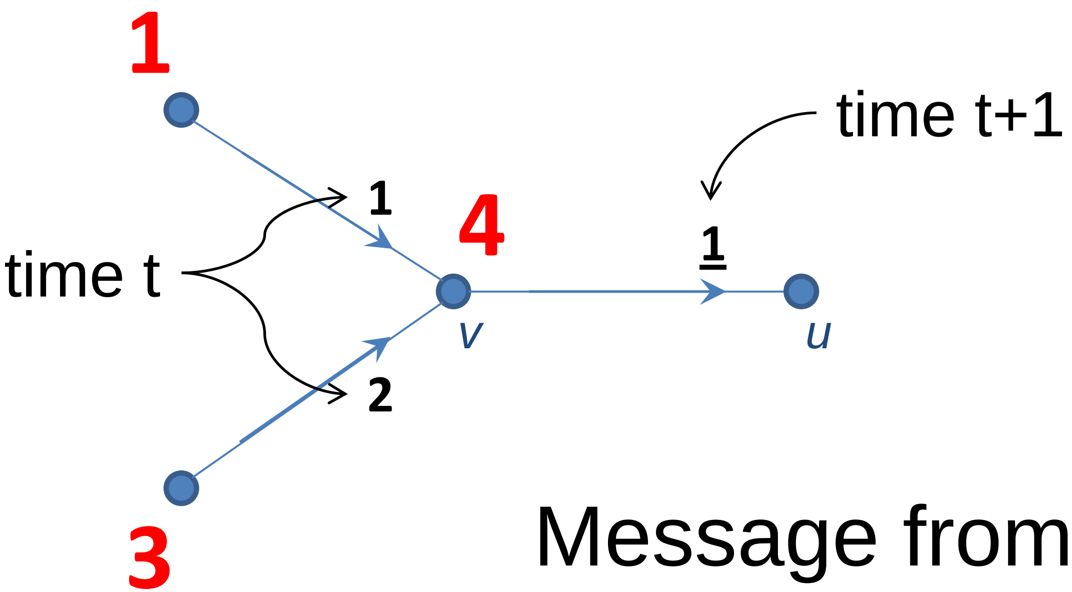
Greedy algorithm on trees



Greedy algorithm on trees



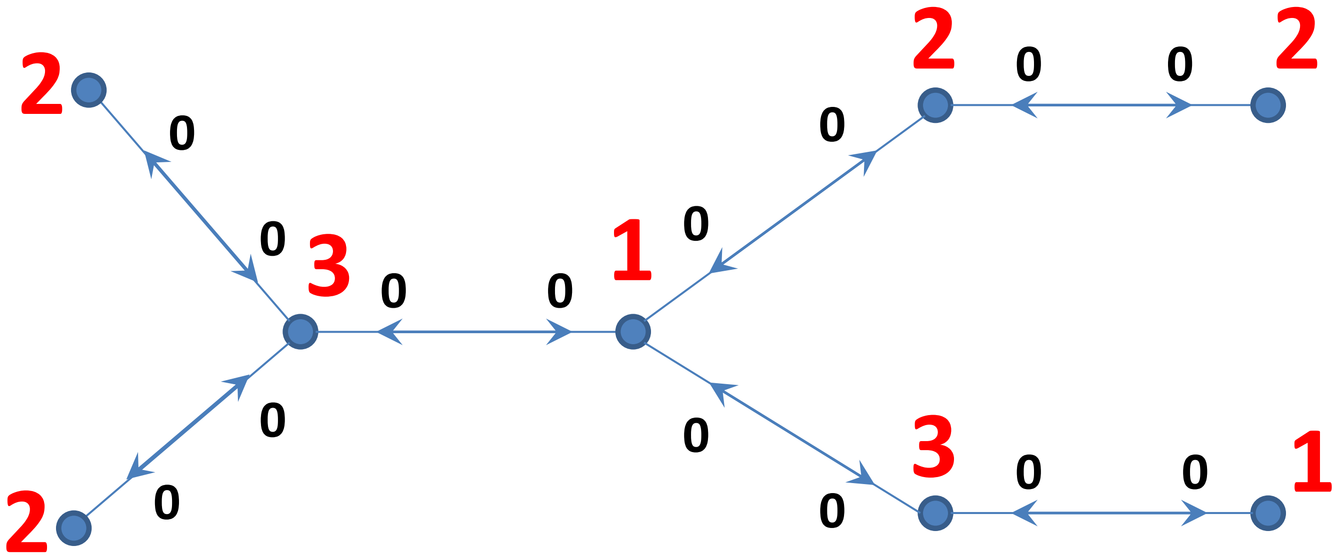
Message-passing version: BP at 0 temperature



Message from *v* to *u*:

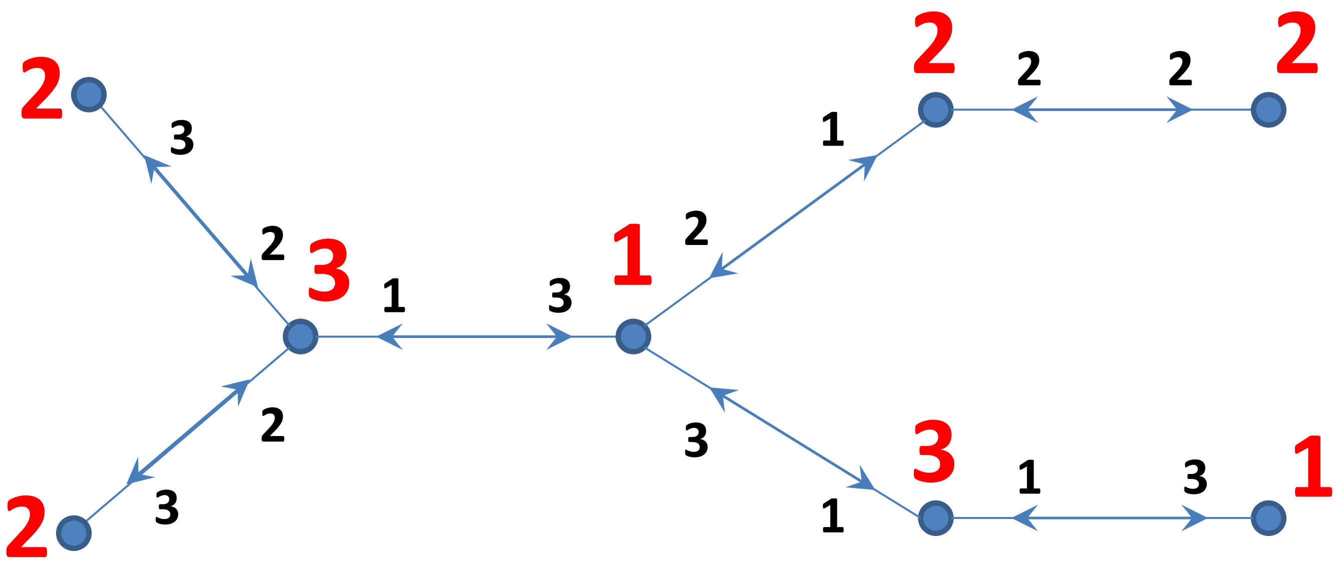
$$I_{v\to u}^{t+1} = \left(b_v - \sum_{w\in\partial v\backslash u} I_{w\to v}^{t}\right)^{+} = \mathrm{P}_{v\to u}\left(\boldsymbol{I}^{t}\right)$$

BP at 0 temperature:
initial messages



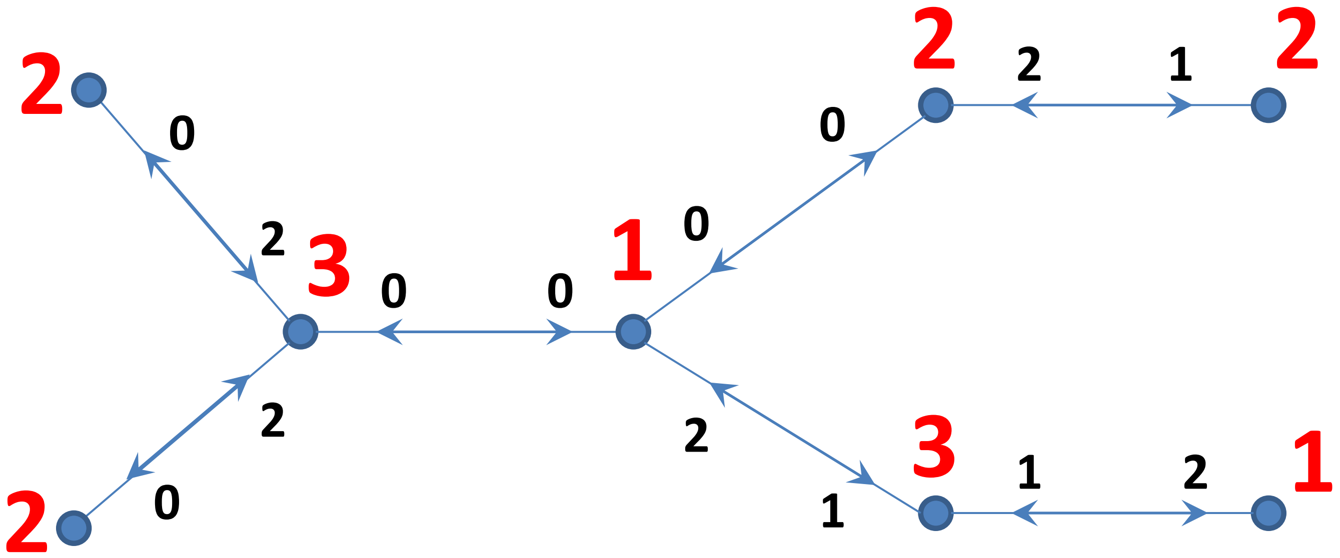
$I_0=0$

BP at 0 temperature:
after 1 iteration



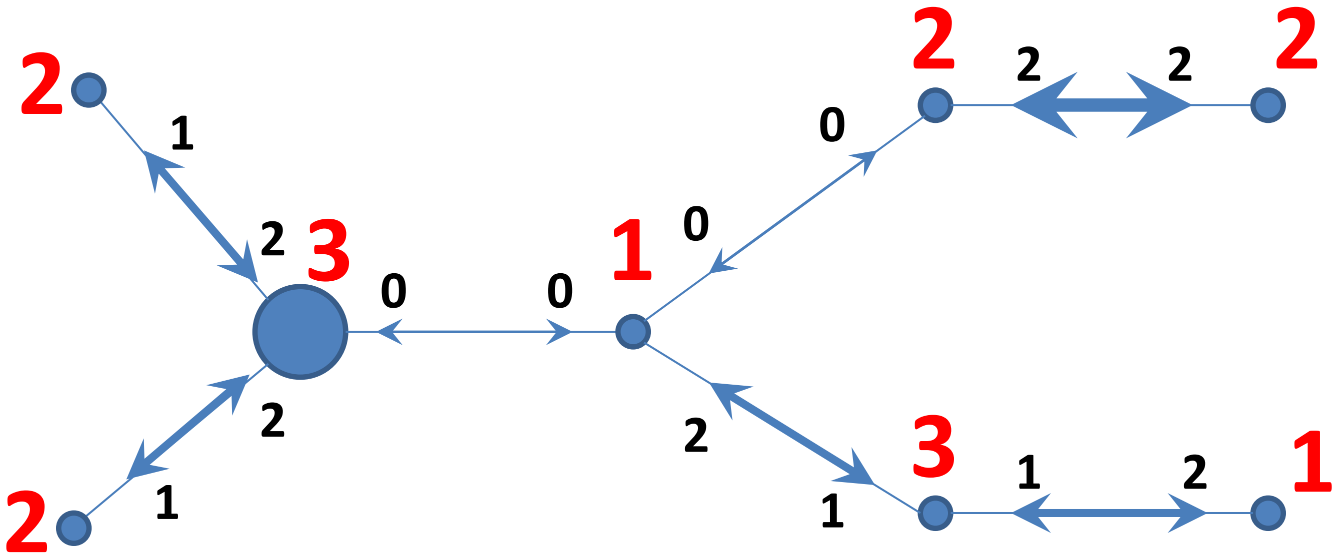
$$\boldsymbol{I}_1 = \mathrm{P}_G(\boldsymbol{I}_0)$$

BP at 0 temperature:
keep iterating...



$$I_2 = P_G(I_1)$$

BP at 0 temperature:
fixed-point



$$\boldsymbol{l}_3 = P_G(\boldsymbol{l}_2) = P_G(\boldsymbol{l}_3) \Rightarrow \boldsymbol{l}^* = \boldsymbol{l}_3$$

Problem solved?

- On trees, can recover the size of a maximum b-matching from $\boldsymbol{I}^*$
- However, we are interested in sequences of graphs that are asymptotically tree-like
  - Message passing may not converge on those
  - Possibly many fixed-points on an infinite tree
    - Some of them will not yield a maximum b-matching…
- BP at positive temperature T and then $T \to 0$

BP at positive temperature

- Gibbs measure on set of b-matchings

$$\mu_G^{\lambda}(\boldsymbol{\sigma}) = \frac{1}{Z_G(\lambda)} \lambda^{\sum_e \sigma_e} \prod_{v \in V} \mathbb{1}\left(\sum_{e \sim v} \sigma_e\right)$$

- At positive temperature, **BP messages are** **distributions over the integers**
  - Initialization: $m_{v \to u}^{(0)} = (1, 0, \dots)$
  - BP update: $m_{v \to u}^{(k+1)}(x) = \mathrm{R}_{v \to u}^{\lambda}\left[\boldsymbol{m}^{(k)}\right](x)$

$$\mathrm{R}_{v \to u}^{\lambda}[\boldsymbol{m}](x) = C \lambda^{x} \sum_{\boldsymbol{\sigma} \in \mathbb{N}^{\partial v \setminus u}, |\boldsymbol{\sigma}| \le b_v - x} \prod_{w \in \partial v \setminus u} m_{w \to v}(\sigma_{vw})$$

where *C* is a normalization factor

BP estimate

- Define $\mathrm{D}_v\left(\boldsymbol{m}^{\lambda}\right)$ =BP estimate of total usage of edges adjacent to *v*, where $\boldsymbol{m}^{\lambda}$ fixed-point of BP at temperature λ

$$\mathrm{D}_v\left(\boldsymbol{m}^{\lambda}\right)=\frac{\sum_{\boldsymbol{\sigma}\in \mathrm{N}^{\partial v},|\boldsymbol{\sigma}|\le b_v} |\boldsymbol{\sigma}|\prod_{u\in\partial v} m_{u\to v}^{\lambda}(\sigma_{uv})}{\sum_{\boldsymbol{\sigma}\in \mathrm{N}^{\partial v},|\boldsymbol{\sigma}|\le b_v} \prod_{u\in\partial v} m_{u\to v}^{\lambda}(\sigma_{uv})}$$

- $\mathrm{D}_v\left(\boldsymbol{m}^{\lambda}\right)=\mu_G^{\lambda}\left(\sum_{e\sim v}\sigma_e\right)$ when G is a finite tree

Global plan

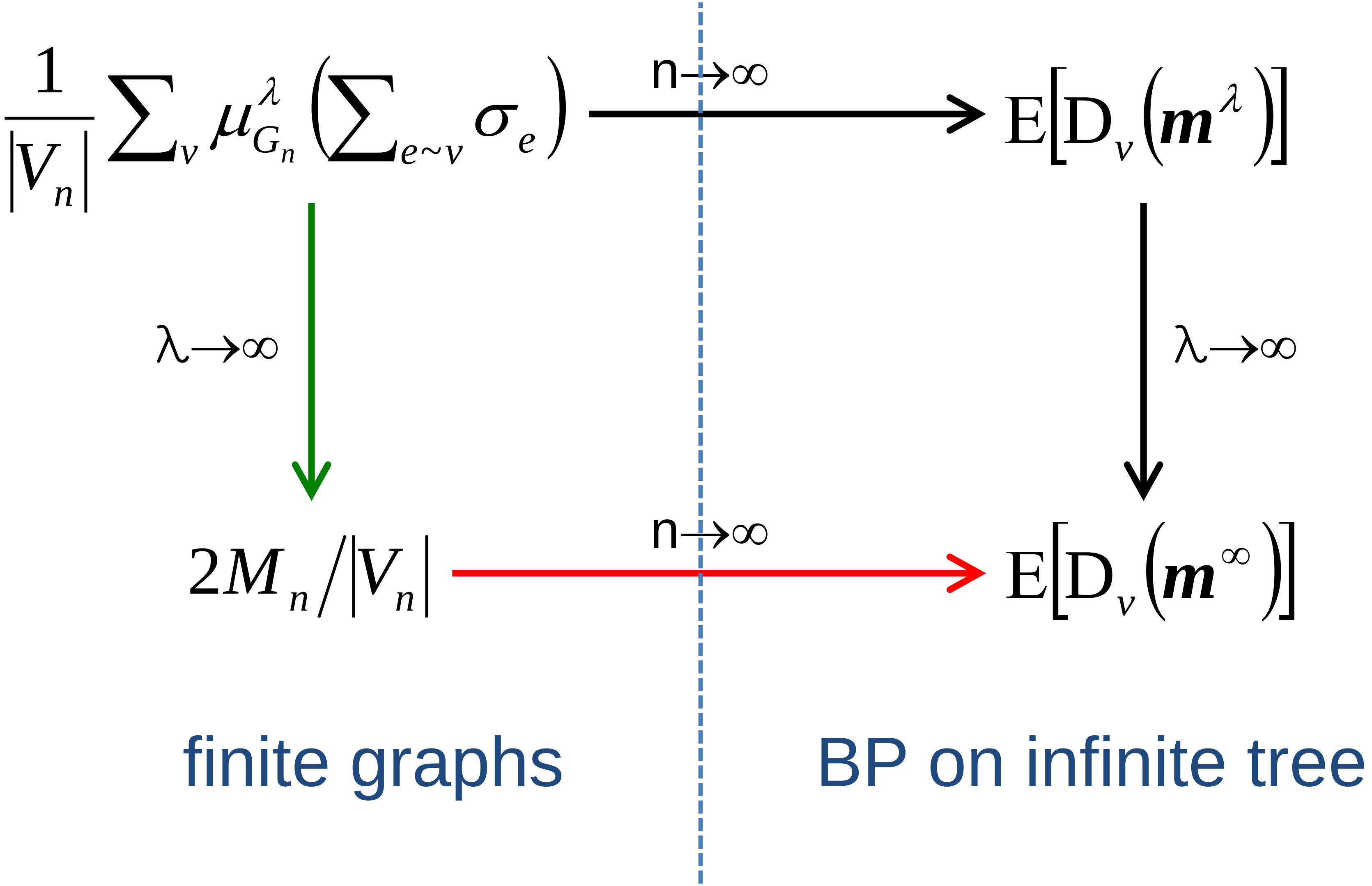
$$\frac{1}{|V_n|}\sum_v \mu_{G_n}^{\lambda}\left(\sum_{e\sim v}\sigma_e\right) \xrightarrow{n\to\infty} \mathrm{E}\left[\mathrm{D}_v\left(\boldsymbol{m}^{\lambda}\right)\right]$$

$$\frac{1}{|V_n|}\sum_v \mu_{G_n}^{\lambda}\left(\sum_{e\sim v}\sigma_e\right) \xrightarrow{\lambda\to\infty} 2M_n/|V_n|$$

$$\mathrm{E}\left[\mathrm{D}_v\left(\boldsymbol{m}^{\lambda}\right)\right] \xrightarrow{\lambda\to\infty} \mathrm{E}\left[\mathrm{D}_v\left(\boldsymbol{m}^{\infty}\right)\right]$$

$$2M_n/|V_n| \xrightarrow{n\to\infty} \mathrm{E}\left[\mathrm{D}_v\left(\boldsymbol{m}^{\infty}\right)\right]$$

Global plan


$\frac{1}{|V_n|}\sum_v \mu_{G_n}^{\lambda}\left(\sum_{e\sim v}\sigma_e\right)$
n→∞
$\mathrm{E}\left[\mathrm{D}_v\left(\boldsymbol{m}^{\lambda}\right)\right]$
λ→∞
λ→∞
$2M_n/|V_n|$
n→∞
$\mathrm{E}\left[\mathrm{D}_v\left(\boldsymbol{m}^{\infty}\right)\right]$
finite graphs
BP on infinite tree

At the end…an ugly formula

$$\lim_{n\to\infty} \frac{2M_n}{|V_n|} = \inf_{q=g\circ g(q)} \mathrm{F}(q)$$

where $$\mathrm{F}(q) = \mathrm{E}\Bigg[b_v 1(|\partial v| > 0) + b_v \wedge \sum_{i=1}^{|\partial v|} Y_i - b_v \wedge \sum_{i=1}^{|\partial v|} \Big(b_v - \sum\nolimits_{j\neq i} X_i\Big)^+ \Bigg]$$

$X_i \sim q$ and $Y_i \sim \mathrm{E}[\mathrm{P}_{v\to u}(\boldsymbol{X})] \sim g(q)$

At the end…an ugly formula

two-step fixed-point

$$\lim_{n\to\infty} \frac{2M_n}{|V_n|} = \inf_{q=g\circ g(q)} \mathrm{F}(q)$$

capacity constraint + incoming messages

where $\mathrm{F}(q) = \mathrm{E}\left[b_v 1(|\partial v| > 0) + b_v \wedge \sum_{i=1}^{|\partial v|} Y_i - b_v \wedge \sum_{i=1}^{|\partial v|} \left(b_v - \sum_{j\neq i} X_i \right)^+ \right]$

- outgoing messages

$X_i \sim q$ and $Y_i \sim \mathrm{E}[\mathrm{P}_{v\to u}(\boldsymbol{X})] \sim g(q)$

Convergence of BP at T>0

- **Does BP converge?**
  - matching case: messages are distributions on {0,1}
    $\rightarrow$ can be encoded by <u>one</u> real number
    - Negative dependence (Pemantle) $\Rightarrow$ BP update operator non-increasing (Salez) $\Rightarrow$ adjacent sequences $\Rightarrow$ […] $\Rightarrow$ unique fixed-point
- **Is $\mathrm{R}_{v\to u}^{\lambda}$ non-increasing?**
  - For general messages and *st*-order, **no!**
  - Restrict to log-concave distributions
    $\rightarrow$ $m_{v\to u}(x)/m_{v\to u}(x-1)$ non-increasing in $x\in\mathrm{N}^{*}$
    - Totally Positive functions (Karlin) $\Rightarrow$ $\mathrm{R}_{v\to u}^{\lambda}$ **non-increasing for *lr*-order** $\Rightarrow$ […] $\Rightarrow$ unique fixed-point

Merci!

Questions?

Applications: cuckoo hashing

Introduced by Pagh & Rodler, ESA'01

- m balls and n bins (mapping of objects and keys)
- Each ball is proposed 2 bins at random
- How large can m be such that it is possible to put each ball into one of its proposed bin, with no collisions allowed?

- Generalizations:
  - More choices per ball $\rightarrow$ still a matching problem
  - Larger bin capacity $\rightarrow$ b>1 on one side
  - Batches of balls with same choices $\rightarrow$ b>1 on both sides $\Rightarrow$ edges may be used multiple times!

Monotonicity as T→0

$$\begin{array}{ccc}
\frac{1}{|V_n|}\sum_v \mu_{G_n}^{\lambda}\left(\sum_{e\sim v}\sigma_e\right) & \xrightarrow{n\to\infty} & \mathrm{E}\left[\mathrm{D}_v\left(\boldsymbol{m}^{\lambda}\right)\right] \\
\downarrow \lambda\to\infty & & \downarrow \lambda\to\infty \\
2M_n/|V_n| & \xrightarrow{n\to\infty} & \mathrm{E}\left[\mathrm{D}_v\left(\boldsymbol{m}^{\infty}\right)\right]
\end{array}$$

- Limit of $\mu_{G_n}^{\lambda}$ as $\lambda\to\infty$ is the uniform law over the set of maximum b-matchings

Correlation decay at T>0

$$\begin{array}{ccc} \frac{1}{|V_n|}\sum_v \mu^{\lambda}_{G_n}\left(\sum_{e\sim v}\sigma_e\right) & \xrightarrow{\text{n}\to\infty} & \mathrm{E}\left[\mathrm{D}_v\left(\boldsymbol{m}^{\lambda}\right)\right] \\ \Big\downarrow \lambda\to\infty & & \Big\downarrow \lambda\to\infty \\ 2M_n/|V_n| & \xrightarrow{\text{n}\to\infty} & \mathrm{E}\left[\mathrm{D}_v\left(\boldsymbol{m}^{\infty}\right)\right] \end{array}$$

- Unicity of BP fixed-point at T>0
- G_n converges locally weakly to G
- G is a tree, hence BP estimate is correct

Monotonicity + continuity

$$
\begin{array}{ccc}
\frac{1}{|V_n|}\sum_v \mu_{G_n}^{\lambda}\left(\sum_{e\sim v}\sigma_e\right) & \xrightarrow{\;n\to\infty\;} & \mathrm{E}\left[\mathrm{D}_v\left(\boldsymbol{m}^{\lambda}\right)\right] \\
\Big\downarrow{\scriptstyle \lambda\to\infty} & & \Big\downarrow{\scriptstyle \lambda\to\infty} \\
2M_n/|V_n| & \xrightarrow{\;n\to\infty\;} & \mathrm{E}\left[\mathrm{D}_v\left(\boldsymbol{m}^{\infty}\right)\right]
\end{array}
$$

- $\boldsymbol{m}^{\infty} = \lim\uparrow_{\lambda\to\infty} \boldsymbol{m}^{\lambda}$
- D_v continuous

Convexity argument

$$
\begin{array}{ccc}
\dfrac{1}{|V_n|}\sum_v \mu_{G_n}^{\lambda}\left(\sum_{e\sim v}\sigma_e\right) & \xrightarrow{\ n\to\infty\ } & \mathrm{E}\left[\mathrm{D}_v\left(\boldsymbol{m}^{\lambda}\right)\right] \\
\Big\downarrow \lambda\to\infty & & \Big\downarrow \lambda\to\infty \\
2M_n/|V_n| & \xrightarrow{\ n\to\infty\ } & \mathrm{E}\left[\mathrm{D}_v\left(\boldsymbol{m}^{\infty}\right)\right]
\end{array}
$$

- Uniform control in *n*
- Entropy term in the free energy becomes negligeable as $T\to 0$

Non-increasing operator?

Ex: b_v=3, $\partial v = \{u_1, u_2, u_3\}$, $m_{u_1 \to v} = \left(\frac{1}{2}, \frac{1}{2}, 0, 0\right) = m'_{u_1 \to v}$

and $m_{u_2 \to v} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{3}, \frac{1}{6}\right) <_{st} m'_{u_2 \to v} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{6}, \frac{1}{3}\right)$

$$\mathrm{R}^1_{v \to u_3}[\boldsymbol{m}] = \frac{12}{25}\left(\frac{11}{12}, \frac{8}{12}, \frac{3}{8}, \frac{1}{8}\right) <_{st} \mathrm{R}^1_{v \to u_3}[\boldsymbol{m}'] = \frac{12}{23}\left(\frac{10}{12}, \frac{7}{12}, \frac{3}{8}, \frac{1}{8}\right)$$

$\Rightarrow$ **no monotonicity for general messages**