Load Balancing in Large Graphs

Venkat Anantharam

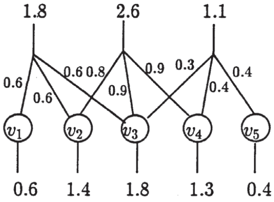
EECS Department
University of California, Berkeley

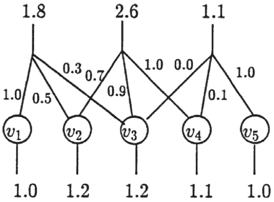
November 26, 2014

Institute for Network Coding
CUHK, Hong Kong
(Joint work with Justin Salez)

Resource allocation

Consumers above, Resources below





Balanced resource allocation

Balanced resource allocation

- Let f be a convex function on the nonnegative reals.

Balanced resource allocation

- Let f be a convex function on the nonnegative reals.
- Given an assignment θ, let the objective be to minimize

$$J(\theta) := \sum_{i=1}^{M} f(\partial\theta(i)) \ .$$

where $\partial\theta(i)$ is the load at resource i and M is the number of resources.

Balanced resource allocation

- Let f be a convex function on the nonnegative reals.
- Given an assignment θ, let the objective be to minimize

$$J(\theta) := \sum_{i=1}^{M} f(\partial\theta(i)) \ .$$

  where $\partial\theta(i)$ is the load at resource i and M is the number of resources.
- Theorem (*Hajek*): The assignment θ minimizes $J(\theta)$ iff for all pairs of resources i, i' available to consumer u we have $\theta_u(i) = 0$ whenever $\partial\theta(i) > \partial\theta(i')$.

Balanced resource allocation

- Let f be a convex function on the nonnegative reals.
- Given an assignment θ, let the objective be to minimize

$$J(\theta) := \sum_{i=1}^{M} f(\partial\theta(i)) \ .$$

  where $\partial\theta(i)$ is the load at resource i and M is the number of resources.
- Theorem (*Hajek*): The assignment θ minimizes $J(\theta)$ iff for all pairs of resources i, i' available to consumer u we have $\theta_u(i) = 0$ whenever $\partial\theta(i) > \partial\theta(i')$.
- Note that the condition for an assignment to be balanced does not depend on f.

Our goals

Our goals

- We want to understand the local environment of a typical agent (consumer, resource) in resource allocation problem with many agents.

Our goals

- We want to understand the local environment of a typical agent (consumer, resource) in resource allocation problem with many agents.
- More ambitiously, we would like to be able to handle cost structures depending on the resource types and/or cost structures depending on the consumer types and the allocation pattern given to the consumer.

Our goals

- We want to understand the local environment of a typical agent (consumer, resource) in resource allocation problem with many agents.
- More ambitiously, we would like to be able to handle cost structures depending on the resource types and/or cost structures depending on the consumer types and the allocation pattern given to the consumer.
- Even more ambitiously, we would like to be able to handle game theoretic formulations where the consumers and/or resources are selfish optimizers.

Our goals

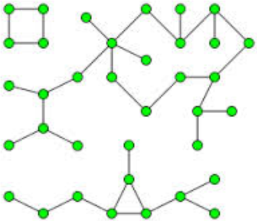
- We want to understand the local environment of a typical agent (consumer, resource) in resource allocation problem with many agents.
- More ambitiously, we would like to be able to handle cost structures depending on the resource types and/or cost structures depending on the consumer types and the allocation pattern given to the consumer.
- Even more ambitiously, we would like to be able to handle game theoretic formulations where the consumers and/or resources are selfish optimizers.
- We cannot do any of this at this stage.

Our goals

- We want to understand the local environment of a typical agent (consumer, resource) in resource allocation problem with many agents.
- More ambitiously, we would like to be able to handle cost structures depending on the resource types and/or cost structures depending on the consumer types and the allocation pattern given to the consumer.
- Even more ambitiously, we would like to be able to handle game theoretic formulations where the consumers and/or resources are selfish optimizers.
- We cannot do any of this at this stage.
- What we can do is to understand the local structure of the basic load balancing problem in the case of large sparse graphs .

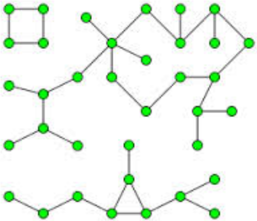
Graphs

Graphs



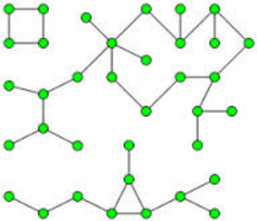
- A graph corresponds to a load balancing problem where each consumer has access to two resources.

Graphs



- A graph corresponds to a load balancing problem where each consumer has access to two resources.
- Each edge is a consumer with one unit of load and has to decide how to distribute its load between the two vertices that define the edge.

Graphs



- A graph corresponds to a load balancing problem where each consumer has access to two resources.
- Each edge is a consumer with one unit of load and has to decide how to distribute its load between the two vertices that define the edge.
- Multiple edges between a pair of vertices are okay.

Load percolation

- The problem is challenging because the local structure of the balanced allocation depends on the global structure of the graph, not just on its local structure.

Load percolation

- The problem is challenging because the local structure of the balanced allocation depends on the global structure of the graph, not just on its local structure.

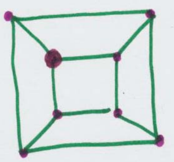


Figure: Graph A

Load percolation

- The problem is challenging because the local structure of the balanced allocation depends on the global structure of the graph, not just on its local structure.

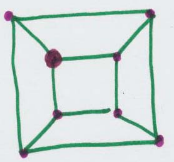


Figure: Graph A

Figure: Graph B

Load percolation

- The problem is challenging because the local structure of the balanced allocation depends on the global structure of the graph, not just on its local structure.

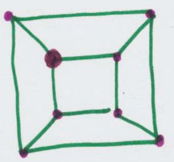


Figure: Graph A

Figure: Graph B

- The marked vertex in graph A has the same depth-1 neighborhood as the root in graph B .

Load percolation

- The problem is challenging because the local structure of the balanced allocation depends on the global structure of the graph, not just on its local structure.

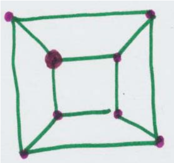


Figure: Graph A

Figure: Graph B

- The marked vertex in graph A has the same depth-1 neighborhood as the root in graph B .
- However the induced balanced load is $\frac{3}{2}$ at each vertex in graph A and is $\frac{4}{5}$ in graph B .

Load percolation

- The problem is challenging because the local structure of the balanced allocation depends on the global structure of the graph, not just on its local structure.

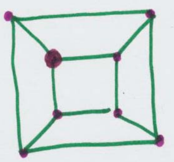


Figure: Graph A

Figure: Graph B

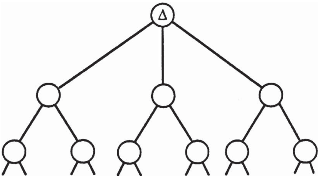
- The marked vertex in graph A has the same depth-1 neighborhood as the root in graph B .
- However the induced balanced load is $\frac{3}{2}$ at each vertex in graph A and is $\frac{4}{5}$ in graph B .
- The phenomenon underlying this is called *load percolation* by Hajek.

Nonuniqueness in the limit

- An infinite sparse graph can exhibit nonuniqueness in its balanced allocations.

Nonuniqueness in the limit

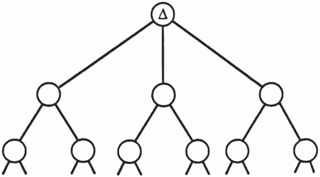
- An infinite sparse graph can exhibit nonuniqueness in its balanced allocations.



- In this infinite 3-regular tree, start by assigning the load of each edge to the vertex that is furthest from the marked vertex.

Nonuniqueness in the limit

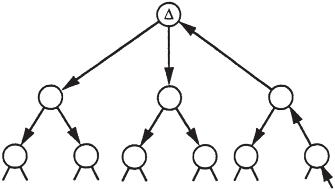
- An infinite sparse graph can exhibit nonuniqueness in its balanced allocations.



- In this infinite 3-regular tree, start by assigning the load of each edge to the vertex that is furthest from the marked vertex.
- This gives induced load 1 at all vertices except for the marked one, which has induced load 0.

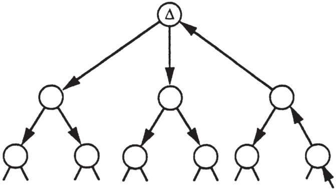
Nonuniqueness in the limit

- Pick a path from infinity to the marked node and flip the allocations of edges along this path.



Nonuniqueness in the limit

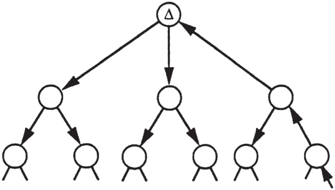
- Pick a path from infinity to the marked node and flip the allocations of edges along this path.



- This allocation is balanced. Each vertex has induced load 1.

Nonuniqueness in the limit

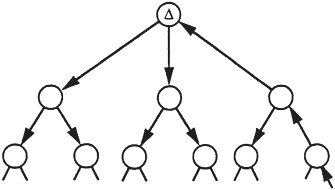
- Pick a path from infinity to the marked node and flip the allocations of edges along this path.



- This allocation is balanced. Each vertex has induced load 1.
- Now flip the allocation of each edge.

Nonuniqueness in the limit

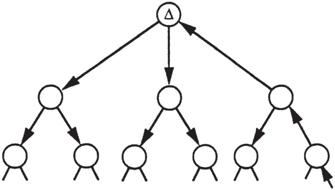
- Pick a path from infinity to the marked node and flip the allocations of edges along this path.



- This allocation is balanced. Each vertex has induced load 1.
- Now flip the allocation of each edge.
- This is another balanced allocation !! . The induced load at each vertex is 2.

Nonuniqueness in the limit

- Pick a path from infinity to the marked node and flip the allocations of edges along this path.



- This allocation is balanced. Each vertex has induced load 1.
- Now flip the allocation of each edge.
- This is another balanced allocation !! . The induced load at each vertex is 2.
- These examples are due to Hajek.

Numerics on Erdös-Rényi graphs (*Hajek*)

αM consumers and M resources; edges picked at random

Numerics on Erdös-Rényi graphs (*Hajek*)

αM consumers and M resources; edges picked at random

SAMPLE LOAD DISTRIBUTION *BEFORE* BALANCING ($\alpha = 2$, $M = 10000$)

τ	Load $\leq \tau$	Load $= \tau$
0.0	201	201
0.5	921	720
1.0	2382	1461
1.5	4299	1917
2.0	6291	1992
2.5	7896	1605
3.0	8899	1003
3.5	9472	573
4.0	9778	306
4.5	9912	134
5.0	9962	50
5.5	9987	25
6.0	10000	13

Numerics on Erdös-Rényi graphs (*Hajek*)

αM consumers and M resources; edges picked at random

Sample Load Distribution *Before* Balancing ($\alpha = 2$, $M = 10000$)

τ	Load $\leq \tau$	Load $= \tau$
0.0	201	201
0.5	921	720
1.0	2382	1461
1.5	4299	1917
2.0	6291	1992
2.5	7896	1605
3.0	8899	1003
3.5	9472	573
4.0	9778	306
4.5	9912	134
5.0	9962	50
5.5	9987	25
6.0	10000	13

Sample Load Distribution *After* Balancing ($\alpha = 2$, $M = 10000$)

τ	Load $\leq \tau$	Load $= \tau$	Product
0.00000000	201	201	0
0.50000000	223	22	11
1.00000000	992	769	769
1.25000000	996	4	5
1.33333333	1023	27	36
1.50000000	1239	216	324
1.60000000	1244	5	8
1.66666667	1313	69	115
1.75000000	1353	40	70
1.77777778	1362	9	16
1.80000000	1392	30	54
1.83333333	1398	6	11
1.85714286	1405	7	13
1.92307692	1418	13	25
2.00000000	3316	1898	3796
2.07692308	3329	13	27
2.11111111	3338	9	19
2.12500000	3362	24	51
2.14285714	3404	42	90
2.16666667	3440	36	78
2.18181818	3462	22	48
2.20000000	3562	100	220
2.20782852	10000	6438	14214

Numerics on Erdös-Rényi graphs (*Hajek*)

SAMPLE LOAD DISTRIBUTION *AFTER* BALANCING ($\alpha = 10$, $M = 10000$)

τ	Load $\leq \tau$	Load $= \tau$	Product
6.00000000	2	2	12
7.00000000	6	4	28
8.00000000	17	11	88
9.00000000	51	34	306
9.33333333	54	3	28
9.50000000	56	2	19
10.00000000	114	58	580
10.00799110	10000	9886	98939

Poisson Galton-Watson tree

- The numerics suggest that there should be a well defined limiting distribution ($M \to \infty$) for the induced load (in a balanced allocation) at a typical vertex.

Poisson Galton-Watson tree

- The numerics suggest that there should be a well defined limiting distribution ($M \to \infty$) for the induced load (in a balanced allocation) at a typical vertex.
- The local environment of a typical vertex in an Erdös - Rényi graph converges to a Poisson Galton-Watson tree as $M \to \infty$.

Poisson Galton-Watson tree

- The numerics suggest that there should be a well defined limiting distribution ($M \to \infty$) for the induced load (in a balanced allocation) at a typical vertex.
- The local environment of a typical vertex in an Erdös - Rényi graph converges to a Poisson Galton-Watson tree as $M \to \infty$.

Poisson Galton-Watson tree : Start with a root, pick a Poisson (λ) number of neighbors (at depth 1). For each of these, independently pick a Poisson (λ) number of neighbors (at depth 2), etc.

Poisson Galton-Watson tree

- The numerics suggest that there should be a well defined limiting distribution ($M \to \infty$) for the induced load (in a balanced allocation) at a typical vertex.
- The local environment of a typical vertex in an Erdös - Rényi graph converges to a Poisson Galton-Watson tree as $M \to \infty$.

Poisson Galton-Watson tree : Start with a root, pick a Poisson (λ) number of neighbors (at depth 1). For each of these, independently pick a Poisson (λ) number of neighbors (at depth 2), etc.

- Natural guess: the limiting induced load distribution obeys a fixed point equation (a recursive distributional equation).

Poisson Galton-Watson tree

- The numerics suggest that there should be a well defined limiting distribution ($M \to \infty$) for the induced load (in a balanced allocation) at a typical vertex.
- The local environment of a typical vertex in an Erdös - Rényi graph converges to a Poisson Galton-Watson tree as $M \to \infty$.

Poisson Galton-Watson tree : Start with a root, pick a Poisson (λ) number of neighbors (at depth 1). For each of these, independently pick a Poisson (λ) number of neighbors (at depth 2), etc.

- Natural guess: the limiting induced load distribution obeys a fixed point equation (a recursive distributional equation).
- This was conjectured by Hajek.

Our contribution

- We verify this conjecture of Hajek as a special case of a broader result.

Our contribution

- We verify this conjecture of Hajek as a special case of a broader result.
- Our results are in the language of local weak convergence of sequences of graphs, also called the objective method .

Our contribution

- We verify this conjecture of Hajek as a special case of a broader result.
- Our results are in the language of local weak convergence of sequences of graphs, also called the objective method .
- In this theory graphs are viewed through the lens of probability distributions on rooted graphs.

Our contribution

- We verify this conjecture of Hajek as a special case of a broader result.
- Our results are in the language of local weak convergence of sequences of graphs, also called the objective method .
- In this theory graphs are viewed through the lens of probability distributions on rooted graphs.
- We prove that there is a uniquely defined balanced allocation associated to any probability distribution on infinite rooted graphs that can arise as a local weak limit of a sequence of finite graphs.

Our contribution

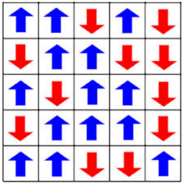
- We verify this conjecture of Hajek as a special case of a broader result.
- Our results are in the language of local weak convergence of sequences of graphs, also called the objective method .
- In this theory graphs are viewed through the lens of probability distributions on rooted graphs.
- We prove that there is a uniquely defined balanced allocation associated to any probability distribution on infinite rooted graphs that can arise as a local weak limit of a sequence of finite graphs.
- The unique balanced allocation on the finite graphs converges to the corresponding unique balanced allocation on its local weak limit.

Our contribution

- We verify this conjecture of Hajek as a special case of a broader result.
- Our results are in the language of local weak convergence of sequences of graphs, also called the objective method .
- In this theory graphs are viewed through the lens of probability distributions on rooted graphs.
- We prove that there is a uniquely defined balanced allocation associated to any probability distribution on infinite rooted graphs that can arise as a local weak limit of a sequence of finite graphs.
- The unique balanced allocation on the finite graphs converges to the corresponding unique balanced allocation on its local weak limit.
- The induced load distribution at the root in the infinite limit rooted graph obeys the expected recursive distributional equation.

Contrasting with mean field methods

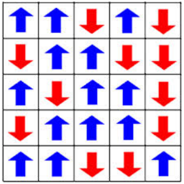
- Consider the dynamics of spin flips on a $\sqrt{n} \times \sqrt{n}$ grid:



Each spin flips at a rate propotional to the number of neighbors of opposite spin.

Contrasting with mean field methods

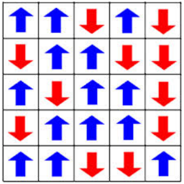
- Consider the dynamics of spin flips on a $\sqrt{n} \times \sqrt{n}$ grid:



  Each spin flips at a rate propotional to the number of neighbors of opposite spin.
- In a mean field model a spin is thought of as picking another spin at random from the remaining $n - 1$ spins and flipping to its orientation; rates normalized.

Contrasting with mean field methods

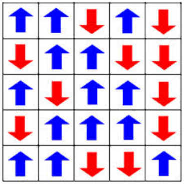
- Consider the dynamics of spin flips on a $\sqrt{n} \times \sqrt{n}$ grid:



  Each spin flips at a rate propotional to the number of neighbors of opposite spin.
- In a mean field model a spin is thought of as picking another spin at random from the remaining $n - 1$ spins and flipping to its orientation; rates normalized.
- The limit mean field model views a single spin as interacting with a time-varying $[0, 1]$-valued variable representing overall average orientation, evolving like individual spins.

Contrasting with mean field methods

- Consider the dynamics of spin flips on a $\sqrt{n} \times \sqrt{n}$ grid:



  Each spin flips at a rate propotional to the number of neighbors of opposite spin.
- In a mean field model a spin is thought of as picking another spin at random from the remaining $n - 1$ spins and flipping to its orientation; rates normalized.
- The limit mean field model views a single spin as interacting with a time-varying $[0, 1]$-valued variable representing overall average orientation, evolving like individual spins.
- The objective method limit views a single spin as the spin at the origin in an infinite grid of spins.

The objective method

- $\mathcal{G}_*$ denotes the set of locally finite connected rooted graphs considered up to rooted isomorphism.

The objective method

- $\mathcal{G}_*$ denotes the set of locally finite connected rooted graphs considered up to rooted isomorphism.
- The distance between two elements of $\mathcal{G}_*$ is $\frac{1}{1+r}$, where r is the largest depth of a neighborhood around the root up to which they agree.

The objective method

- $\mathcal{G}_*$ denotes the set of locally finite connected rooted graphs considered up to rooted isomorphism.
- The distance between two elements of $\mathcal{G}_*$ is $\frac{1}{1+r}$, where r is the largest depth of a neighborhood around the root up to which they agree.
- This distance makes $\mathcal{G}_*$ into a complete separable metric space.

The objective method

- $\mathcal{G}_*$ denotes the set of locally finite connected rooted graphs considered up to rooted isomorphism.
- The distance between two elements of $\mathcal{G}_*$ is $\frac{1}{1+r}$, where r is the largest depth of a neighborhood around the root up to which they agree.
- This distance makes $\mathcal{G}_*$ into a complete separable metric space.
- A fixed finite graph G corresponds to a probability distribution on $\mathcal{G}_*$ by picking the root at random from the vertices of G.

The objective method

- $\mathcal{G}_*$ denotes the set of locally finite connected rooted graphs considered up to rooted isomorphism.
- The distance between two elements of $\mathcal{G}_*$ is $\frac{1}{1+r}$, where r is the largest depth of a neighborhood around the root up to which they agree.
- This distance makes $\mathcal{G}_*$ into a complete separable metric space.
- A fixed finite graph G corresponds to a probability distribution on $\mathcal{G}_*$ by picking the root at random from the vertices of G.
- A sequence of finite graphs is said to converge in the sense of local weak convergence if the corresponding probability distributions on $\mathcal{G}_*$ converge weakly.

The objective method

- $\mathcal{G}_*$ denotes the set of locally finite connected rooted graphs considered up to rooted isomorphism.
- The distance between two elements of $\mathcal{G}_*$ is $\frac{1}{1+r}$, where r is the largest depth of a neighborhood around the root up to which they agree.
- This distance makes $\mathcal{G}_*$ into a complete separable metric space.
- A fixed finite graph G corresponds to a probability distribution on $\mathcal{G}_*$ by picking the root at random from the vertices of G.
- A sequence of finite graphs is said to converge in the sense of local weak convergence if the corresponding probability distributions on $\mathcal{G}_*$ converge weakly.

The definitions extend naturally to marked graphs , i.e. graphs where each edge carries an element of some other separable metric space.

The objective method

- $\mathcal{G}_{**}$ denotes the set of locally finite connected graphs with a distinguished oriented edge, considered up to isomorphism (preserving the distinguished oriented edge).

The objective method

- $\mathcal{G}_{**}$ denotes the set of locally finite connected graphs with a distinguished oriented edge, considered up to isomorphism (preserving the distinguished oriented edge).
- $\mathcal{G}_{**}$ can be metrized to give a complete separable metric space, just as for $\mathcal{G}_{*}$.

The objective method

- $\mathcal{G}_{**}$ denotes the set of locally finite connected graphs with a distinguished oriented edge, considered up to isomorphism (preserving the distinguished oriented edge).
- $\mathcal{G}_{**}$ can be metrized to give a complete separable metric space, just as for $\mathcal{G}_{*}$.
- A function $f \ : \ \mathcal{G}_{**} \mapsto \mathbb{R}$ gives rise to a function $\partial f \ : \ \mathcal{G}_{*} \mapsto \mathbb{R}$ via

$$\partial f(G, o) = \sum_{i \sim o} f(G, i, o) \ .$$

The objective method

- $\mathcal{G}_{**}$ denotes the set of locally finite connected graphs with a distinguished oriented edge, considered up to isomorphism (preserving the distinguished oriented edge).
- $\mathcal{G}_{**}$ can be metrized to give a complete separable metric space, just as for $\mathcal{G}_*$.
- A function $f : \mathcal{G}_{**} \mapsto \mathbb{R}$ gives rise to a function $\partial f : \mathcal{G}_* \mapsto \mathbb{R}$ via

$$\partial f(G, o) = \sum_{i \sim o} f(G, i, o) \ .$$

- A probability distribution μ on $\mathcal{G}_*$ gives rise to a measure $\vec{\mu}$ on $\mathcal{G}_{**}$ via

$$\int_{\mathcal{G}_{**}} f d\vec{\mu} = \int_{\mathcal{G}_*} \partial f d\mu \ , \quad \text{for all bounded continuous } f \ .$$

The objective method

- $\mathcal{G}_{**}$ denotes the set of locally finite connected graphs with a distinguished oriented edge, considered up to isomorphism (preserving the distinguished oriented edge).
- $\mathcal{G}_{**}$ can be metrized to give a complete separable metric space, just as for $\mathcal{G}_*$.
- A function $f : \mathcal{G}_{**} \mapsto \mathbb{R}$ gives rise to a function $\partial f : \mathcal{G}_* \mapsto \mathbb{R}$ via

$$\partial f(G, o) = \sum_{i \sim o} f(G, i, o) .$$

- A probability distribution μ on $\mathcal{G}_*$ gives rise to a measure $\vec{\mu}$ on $\mathcal{G}_{**}$ via

$$\int_{\mathcal{G}_{**}} f d\vec{\mu} = \int_{\mathcal{G}_*} \partial f d\mu \ , \quad \text{for all bounded continuous } f \ .$$

- Note that $\vec{\mu}(\mathcal{G}_{**}) = \deg(\mu) := \int_{\mathcal{G}_*} \deg(\text{root}) d\mu$.

Unimodularity

- Given $f : \mathcal{G}_{**} \mapsto \mathbb{R}$, define $f^* : \mathcal{G}_{**} \mapsto \mathbb{R}$ via

$$f^*(G, i, o) = f(G, o, i) \ .$$

Unimodularity

- Given $f \ : \ \mathcal{G}_{**} \mapsto \mathbb{R}$, define $f^* \ : \ \mathcal{G}_{**} \mapsto \mathbb{R}$ via

$$f^*(G, i, o) = f(G, o, i) \ .$$

- A probability distribution μ on $\mathcal{G}_*$ is called unimodular if

$$\int_{\mathcal{G}_{**}} f d\vec{\mu} = \int_{\mathcal{G}_{**}} f^* d\vec{\mu} \ , \quad \text{for all bounded continuous } f \ .$$

Unimodularity

- Given $f \ : \ \mathcal{G}_{**} \mapsto \mathbb{R}$, define $f^* \ : \ \mathcal{G}_{**} \mapsto \mathbb{R}$ via

$$f^*(G, i, o) = f(G, o, i) \ .$$

- A probability distribution μ on $\mathcal{G}_*$ is called unimodular if

$$\int_{\mathcal{G}_{**}} f d\vec{\mu} = \int_{\mathcal{G}_{**}} f^* d\vec{\mu} \ , \quad \text{for all bounded continuous } f \ .$$

- It is known that the local weak limit of any sequence of finite graphs is unimodular (*Aldous and Lyons*).

Our main result

- A function $\Theta : \mathcal{G}_{**} \mapsto [0, 1]$ is called an allocation if $\Theta + \Theta^* = 1$.

Our main result

- A function $\Theta \ : \ \mathcal{G}_{**} \mapsto [0,1]$ is called an allocation if $\Theta + \Theta^* = 1$.
- An allocation Θ is called a balanced allocation for a given unimodular μ if for $\vec{\mu}$ almost all $(G, i.o)$ it holds that

$$\partial\Theta(G,i) < \partial\Theta(G,o) \Longrightarrow \Theta(G,i,o) = 0 \ .$$

Our main result

- A function $\Theta : \mathcal{G}_{**} \mapsto [0,1]$ is called an allocation if $\Theta + \Theta^* = 1$.
- An allocation Θ is called a balanced allocation for a given unimodular μ if for $\vec{\mu}$ almost all $(G, i.o)$ it holds that

$$\partial\Theta(G,i) < \partial\Theta(G,o) \Longrightarrow \Theta(G,i,o) = 0 \ .$$

- We prove that for any unimodular μ with $\deg(\mu) < \infty$ there is a Θ_0 that is a balanced allocation for μ with the property that it simultaneously minimizes $\int_{\mathcal{G}_*} f(\partial\Theta) d\mu$ over allocations Θ for every convex real valued function f on $\mathbb{R}_+$.

Our main result

- A function $\Theta \ : \ \mathcal{G}_{**} \mapsto [0,1]$ is called an allocation if $\Theta + \Theta^* = 1$.
- An allocation Θ is called a balanced allocation for a given unimodular μ if for $\vec{\mu}$ almost all $(G, i.o)$ it holds that

$$\partial\Theta(G,i) < \partial\Theta(G,o) \Longrightarrow \Theta(G,i,o) = 0 \ .$$

- We prove that for any unimodular μ with $\deg(\mu) < \infty$ there is a Θ_0 that is a balanced allocation for μ with the property that it simultaneously minimizes $\int_{\mathcal{G}_*} f(\partial\Theta) d\mu$ over allocations Θ for every convex real valued function f on $\mathbb{R}_+$.
- Further, Θ_0 is μ-almost surely unique.

Our main result

- A function $\Theta : \mathcal{G}_{**} \mapsto [0,1]$ is called an allocation if $\Theta + \Theta^* = 1$.
- An allocation Θ is called a balanced allocation for a given unimodular μ if for $\vec{\mu}$ almost all $(G, i.o)$ it holds that

$$\partial\Theta(G,i) < \partial\Theta(G,o) \Longrightarrow \Theta(G,i,o) = 0 \ .$$

- We prove that for any unimodular μ with $\deg(\mu) < \infty$ there is a Θ_0 that is a balanced allocation for μ with the property that it simultaneously minimizes $\int_{\mathcal{G}_*} f(\partial\Theta) d\mu$ over allocations Θ for every convex real valued function f on $\mathbb{R}_+$.
- Further, Θ_0 is μ-almost surely unique.
- For any sequence of finite graphs with local weak limit μ, the empiricial distribution of the induced load in the unique balanced allocation on these graphs converges weakly to the law of $\partial\Theta_0$ (for the Θ_0 of the limit).

Variational characterization of the limit

- Given unimodular μ on $\mathcal{G}_*$ with $\deg(\mu) < \infty$, define, for each $t \geq 0$,

$$\Phi_\mu(t) := \int_{\mathcal{G}_*} (\partial\Theta_0 - t)^+ d\mu \ .$$

Variational characterization of the limit

- Given unimodular μ on $\mathcal{G}_*$ with $\deg(\mu) < \infty$, define, for each $t \geq 0$,

$$\Phi_\mu(t) := \int_{\mathcal{G}_*} (\partial\Theta_0 - t)^+ d\mu \ .$$

- $t \mapsto \Phi_\mu(t)$ is the mean-excess function of the almost surely unique balanced allocation associated to μ.

Variational characterization of the limit

- Given unimodular μ on $\mathcal{G}_*$ with $\deg(\mu) < \infty$, define, for each $t \geq 0$,

$$\Phi_\mu(t) := \int_{\mathcal{G}_*} (\partial\Theta_0 - t)^+ d\mu \ .$$

- $t \mapsto \Phi_\mu(t)$ is the mean-excess function of the almost surely unique balanced allocation associated to μ.
- We have the variational characterization

$$\Phi_\mu(t) = \max_{f \,:\, \mathcal{G}_* \to [0,1], \text{Borel}} \left\{ \frac{1}{2} \int_{\mathcal{G}_{**}} \hat{f} d\vec{\mu} - t \int_{\mathcal{G}_*} f d\mu \right\} ,$$

for each t, where

$$\hat{f}(G,i,o) := f(G,i) \wedge f(G,o) \ .$$

Intuition behind the variational characterization

- The optimizing function is $f = 1(\partial\Theta_0 > t)$.

Intuition behind the variational characterization

- The optimizing function is $f = 1(\partial\Theta_0 > t)$.
- To check this, observe that

$$
\begin{aligned}
\frac{1}{2}\int_{\mathcal{G}_{**}} \hat{f}\, d\vec{\mu} &= \frac{1}{2}\int_{\mathcal{G}_*} (\partial \hat{f}) d\mu \\
&= \frac{1}{2}\int_{\mathcal{G}_*} \sum_{i\sim o} 1(\partial\Theta_0(G,i) > t \text{ and } \partial\Theta_0(G,o) > t) d\mu
\end{aligned}
$$

Intuition behind the variational characterization

- The optimizing function is $f = 1(\partial\Theta_0 > t)$.
- To check this, observe that

$$\begin{aligned}\frac{1}{2}\int_{\mathcal{G}_{**}} \hat{f}\, d\vec{\mu} &= \frac{1}{2}\int_{\mathcal{G}_*} (\partial \hat{f})\, d\mu \\ &= \frac{1}{2}\int_{\mathcal{G}_*} \sum_{i\sim o} 1(\partial\Theta_0(G,i) > t \text{ and } \partial\Theta_0(G,o) > t)\, d\mu\end{aligned}$$

- Thus

$$\int_{\mathcal{G}_*} (\partial\Theta_0 - t)^+ d\mu = \frac{1}{2}\int_{\mathcal{G}_{**}} \hat{f}\, d\vec{\mu} - t\int_{\mathcal{G}_*} f d\mu\ ,$$

for this choice of f.

Unimodular Galton-Watson trees

- Given a probability distribution $\{\pi(i)\ , \quad i \geq 0\}$ on the nonnegative integers, with finite mean $\sum_i i\pi(i)$, define
$$\hat{\pi}(i) := \frac{(i+1)\pi(i+1)}{\sum_i i\pi(i)}\ , \quad i \geq 0\ .$$
$\{\hat{\pi}(i)\ , \quad i \geq 0\}$ is also a probability distribution.

Unimodular Galton-Watson trees

- Given a probability distribution $\{\pi(i) \ , \quad i \geq 0\}$ on the nonnegative integers, with finite mean $\sum_i i\pi(i)$, define
$$\hat{\pi}(i) := \frac{(i+1)\pi(i+1)}{\sum_i i\pi(i)} \ , \quad i \geq 0 \ .$$
$\{\hat{\pi}(i) \ , \quad i \geq 0\}$ is also a probability distribution.
- The unimodular Galton-Watson tree, UGWT(π) is the random tree constructed as follows: Start with a root and give it a random number of children (at depth 1) with the number of children distributed as π. For each child, give it a random number of children (at depth 2), the number distributed as $\hat{\pi}$, independently. Repeat (using $\hat{\pi}$ from now on).

Unimodular Galton-Watson trees

- Given a probability distribution $\{\pi(i)\ ,\quad i \geq 0\}$ on the nonnegative integers, with finite mean $\sum_i i\pi(i)$, define

$$\hat{\pi}(i) := \frac{(i+1)\pi(i+1)}{\sum_i i\pi(i)}\ ,\quad i \geq 0\ .$$

  $\{\hat{\pi}(i)\ ,\quad i \geq 0\}$ is also a probability distribution.
- The unimodular Galton-Watson tree, UGWT(π) is the random tree constructed as follows: Start with a root and give it a random number of children (at depth 1) with the number of children distributed as π. For each child, give it a random number of children (at depth 2), the number distributed as $\hat{\pi}$, independently. Repeat (using $\hat{\pi}$ from now on).
- Many standard sequences of bipartite graph models, such as the pairing model based on half edges and fixed degree distributions which shows up in the theory of LDPC codes, have a unimodular Galton-Watson tree as their local weak limit

Recursive distributional equation characterization of the limit on unimodular Galton-Watson trees

- If μ is the law of UGWT(π), then for every t, we have

$$\Phi_\mu(t) = \max_{Q=F_{\pi,t}(Q)} \{ \frac{E[D]}{2} P(\xi_1 + \xi_2 > 1) - tP(\xi_1 + \ldots + \xi_D > t) \} \ ,$$

where $F_{\pi,t}(Q)$ is the law of $[1 - t + \xi_1 + \ldots + \xi_{\hat{D}}]_0^1$.

Recursive distributional equation characterization of the limit on unimodular Galton-Watson trees

- If μ is the law of UGWT(π), then for every t, we have

$$\Phi_\mu(t) = \max_{Q=F_{\pi,t}(Q)} \{ \frac{E[D]}{2} P(\xi_1 + \xi_2 > 1) - tP(\xi_1 + \ldots + \xi_D > t) \} ,$$

where $F_{\pi,t}(Q)$ is the law of $[1 - t + \xi_1 + \ldots + \xi_{\hat{D}}]_0^1$.

- Here $[a]_0^1$ equals 0 if $a < 0$, 1 if $a > 1$ and a otherwise. Also, $\hat{D}$ has the law $\hat{\pi}$, D has the law π, and the ξ_i are i.i.d. with law Q.

Recursive distributional equation characterization of the limit on unimodular Galton-Watson trees

- If μ is the law of UGWT(π), then for every t, we have

$$\Phi_\mu(t) = \max_{Q=F_{\pi,t}(Q)} \{\frac{E[D]}{2} P(\xi_1+\xi_2>1) - tP(\xi_1+\ldots+\xi_D>t)\} \ ,$$

  where $F_{\pi,t}(Q)$ is the law of $[1-t+\xi_1+\ldots+\xi_{\hat{D}}]_0^1$.
- Here $[a]_0^1$ equals 0 if $a<0$, 1 if $a>1$ and a otherwise. Also, $\hat{D}$ has the law $\hat{\pi}$, D has the law π, and the ξ_i are i.i.d. with law Q.
- Recall that

$$t \mapsto \Phi_\mu(t) := \int_{\mathcal{G}_*} (\partial\Theta_0 - t)^+ d\mu \ ,$$

  characterizes the limiting distribution of the induced load at the root.

Recursive distributional equation characterization of the limit on unimodular Galton-Watson trees

- If μ is the law of UGWT(π), then for every t, we have

$$\Phi_\mu(t) = \max_{Q=F_{\pi,t}(Q)} \{ \frac{E[D]}{2} P(\xi_1 + \xi_2 > 1) - tP(\xi_1 + \ldots + \xi_D > t) \} ,$$

where $F_{\pi,t}(Q)$ is the law of $[1 - t + \xi_1 + \ldots + \xi_{\hat{D}}]_0^1$.

- Here $[a]_0^1$ equals 0 if $a < 0$, 1 if $a > 1$ and a otherwise. Also, $\hat{D}$ has the law $\hat{\pi}$, D has the law π, and the ξ_i are i.i.d. with law Q.
- Recall that

$$t \mapsto \Phi_\mu(t) := \int_{\mathcal{G}_*} (\partial\Theta_0 - t)^+ d\mu ,$$

characterizes the limiting distribution of the induced load at the root.

- The above recursive distributional characterization of is in effect the one conjectured by Hajek.

Intuition behind the RDE

- We consider the RDE $Q = F_{\pi,t}(Q)$, where $F_{\pi,t}(Q)$ is the law of $[1 - t + \xi_1 + \ldots + \xi_{\hat{D}}]_0^1$, where $\xi_1, \xi_2, \ldots$ are i.i.d with the law Q.

Intuition behind the RDE

- We consider the RDE $Q = F_{\pi,t}(Q)$, where $F_{\pi,t}(Q)$ is the law of $[1 - t + \xi_1 + \ldots + \xi_{\hat{D}}]_0^1$, where $\xi_1, \xi_2, \ldots$ are i.i.d with the law Q.
- Consider an edge (i, o). We are "solving for the load that passes in the direction from o to i.

Intuition behind the RDE

- We consider the RDE $Q = F_{\pi,t}(Q)$, where $F_{\pi,t}(Q)$ is the law of $[1 - t + \xi_1 + \ldots + \xi_{\hat{D}}]_0^1$, where $\xi_1, \xi_2, \ldots$ are i.i.d with the law Q.
- Consider an edge (i, o). We are "solving for the load that passes in the direction from o to i.
- For $1 \leq k \leq \hat{D}$, $1 - \xi_k$ has the meaning of the amount of load that can be absorbed by the k-th child of o (think of i as the parent of o and not as a child), this child of course supporting its own subtree of children, such as to make the net load at that child equal to t.

Intuition behind the RDE

- We consider the RDE $Q = F_{\pi,t}(Q)$, where $F_{\pi,t}(Q)$ is the law of $[1 - t + \xi_1 + \ldots + \xi_{\hat{D}}]_0^1$, where $\xi_1, \xi_2, \ldots$ are i.i.d with the law Q.
- Consider an edge (i, o). We are "solving for the load that passes in the direction from o to i.
- For $1 \leq k \leq \hat{D}$, $1 - \xi_k$ has the meaning of the amount of load that can be absorbed by the k-th child of o (think of i as the parent of o and not as a child), this child of course supporting its own subtree of children, such as to make the net load at that child equal to t.
- The number $[1 - (t - \xi_1 - \ldots - \xi_{\hat{D}})]_0^1$ is then the amount that would be presented in the direction from node o to node i in order to maintain a total load of t at node o.

Convergence of the maximum load

- Under a mild additional on the degree distributions the maximum load also converges to the maximum of the limit.

Convergence of the maximum load

- Under a mild additional on the degree distributions the maximum load also converges to the maximum of the limit.
- This verifies the conjecture of Hajek regarding the limit of the maximum load.

Convergence of the maximum load

- Under a mild additional on the degree distributions the maximum load also converges to the maximum of the limit.
- This verifies the conjecture of Hajek regarding the limit of the maximum load.
- One must exclude “local pockets of high edge density” in the graph.

Convergence of the maximum load

- Under a mild additional on the degree distributions the maximum load also converges to the maximum of the limit.
- This verifies the conjecture of Hajek regarding the limit of the maximum load.
- One must exclude "local pockets of high edge density" in the graph.
- Assume that for some $\lambda > 0$ we have

$$\sup_{n \geq 1} \{ \frac{1}{n} \sum_{i=1}^{n} e^{\lambda d_n(i)} \} < \infty \ .$$

Convergence of the maximum load

- Under a mild additional on the degree distributions the maximum load also converges to the maximum of the limit.
- This verifies the conjecture of Hajek regarding the limit of the maximum load.
- One must exclude "local pockets of high edge density" in the graph.
- Assume that for some $\lambda > 0$ we have

$$\sup_{n \geq 1} \{ \frac{1}{n} \sum_{i=1}^{n} e^{\lambda d_n(i)} \} < \infty \ .$$

- Let $Z_{\delta,t}^{(n)}$ denote the number of subsets S of $\{1, \ldots, n\}$ of size $|S| \leq \delta n$ with edge count $|E(S)| \geq t|S|$ in the given random pairing model. Then we can show that

$$P(Z_{\delta,t}^{(n)} > 0) \to 0 \ , \quad \text{as } n \to \infty \ .$$

This suffices.

Sketch of the proof of the main result

- The key idea is to consider so-called ϵ-balanced allocations, i.e. allocations θ on a locally finite graph G that satisfy

$$\theta(i,j) = \left[\frac{1}{2} + \frac{1}{2\epsilon}(\partial\theta(i) - \partial\theta(j))\right]_0^1 .$$

Sketch of the proof of the main result

- The key idea is to consider so-called ϵ-balanced allocations, i.e. allocations θ on a locally finite graph G that satisfy

$$\theta(i,j) = \left[\frac{1}{2} + \frac{1}{2\epsilon}(\partial\theta(i) - \partial\theta(j))\right]_0^1 .$$

- There is a built-in contractivity in this definition for bounded degree graphs, which allows one to establish the uniqueness of ϵ-balanced allocations for such graphs.

Sketch of the proof of the main result

- The key idea is to consider so-called ϵ-balanced allocations, i.e. allocations θ on a locally finite graph G that satisfy

$$\theta(i,j) = \left[\frac{1}{2} + \frac{1}{2\epsilon}(\partial\theta(i) - \partial\theta(j))\right]_0^1 .$$

- There is a built-in contractivity in this definition for bounded degree graphs, which allows one to establish the uniqueness of ϵ-balanced allocations for such graphs.
- The case of locally finite graphs can be handled by a truncation argument.

Sketch of the proof of the main result

- The key idea is to consider so-called ϵ-balanced allocations, i.e. allocations θ on a locally finite graph G that satisfy

$$\theta(i,j) = \left[\frac{1}{2} + \frac{1}{2\epsilon}(\partial\theta(i) - \partial\theta(j))\right]_0^1 .$$

- There is a built-in contractivity in this definition for bounded degree graphs, which allows one to establish the uniqueness of ϵ-balanced allocations for such graphs.
- The case of locally finite graphs can be handled by a truncation argument.
- The claimed Θ_0 can then be shown to exist as a limit in L^2 of the ϵ-balanced allocations as $\epsilon \to 0$.

Sketch of the proof of the main result

- The key idea is to consider so-called ϵ-balanced allocations, i.e. allocations θ on a locally finite graph G that satisfy

$$\theta(i,j) = \left[\frac{1}{2} + \frac{1}{2\epsilon}(\partial\theta(i) - \partial\theta(j))\right]_0^1 .$$

- There is a built-in contractivity in this definition for bounded degree graphs, which allows one to establish the uniqueness of ϵ-balanced allocations for such graphs.
- The case of locally finite graphs can be handled by a truncation argument.
- The claimed Θ_0 can then be shown to exist as a limit in L^2 of the ϵ-balanced allocations as $\epsilon \to 0$.
- The ϵ-relaxation can be roughly thought of as analogous to working at finite temperature (versus zero temperature) in statistical mechanics.

The End