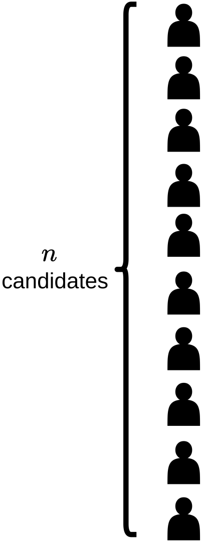
Matching Algorithms in the Sparse Stochastic Block Model

Anna Brandenberger, Byron Chin, Nathan Sheffield and Divya Shyamal

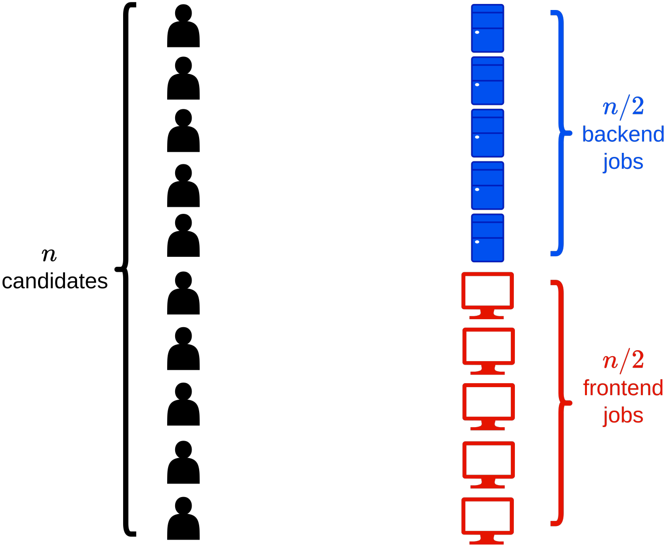
MIT

AofA 2024

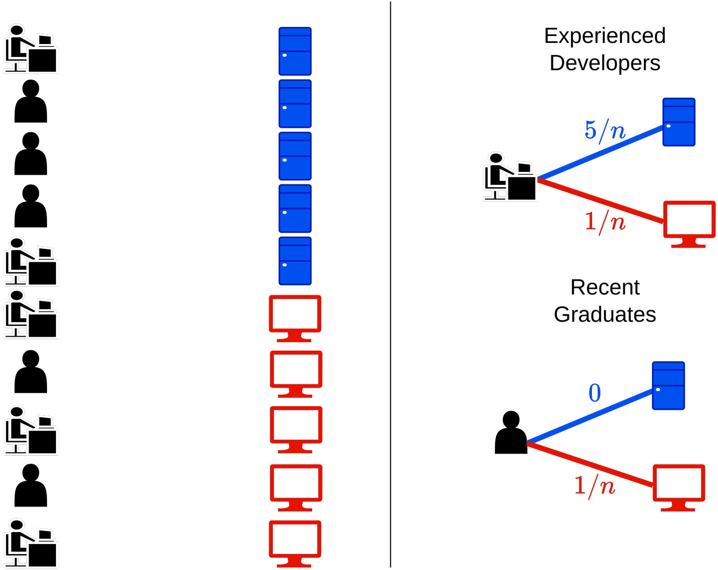
A MOTIVATING FABLE: HIRING A TEAM



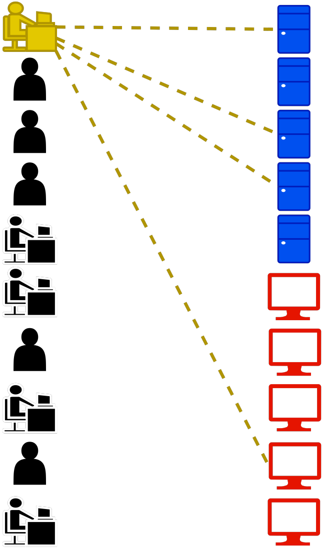
A motivating fable: hiring a team

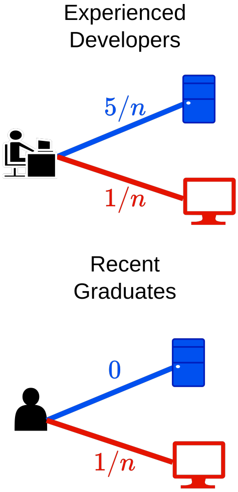


A motivating fable: hiring a team

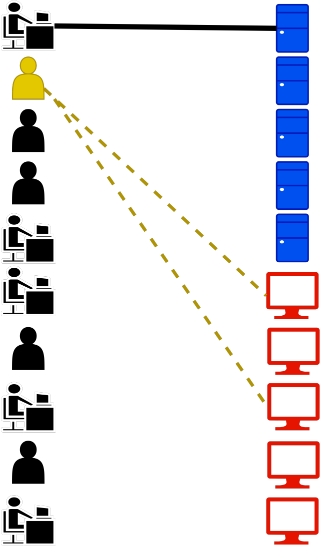


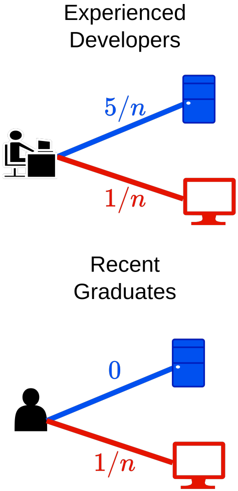
A motivating fable: hiring a team



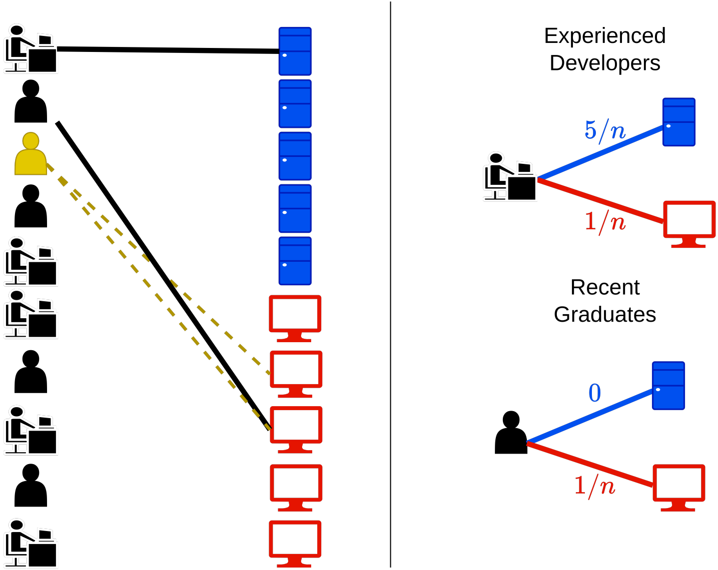


A motivating fable: hiring a team

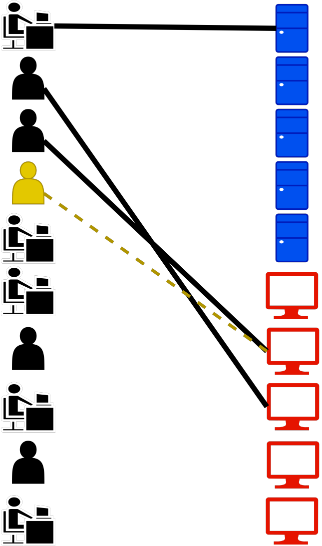


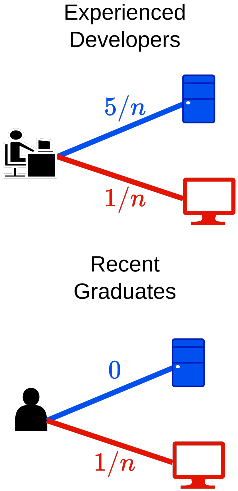


A MOTIVATING FABLE: HIRING A TEAM

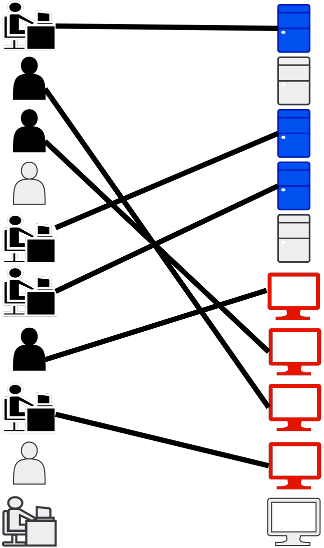


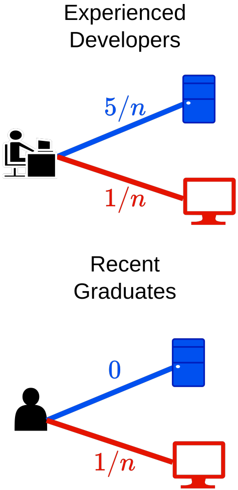
A MOTIVATING FABLE: HIRING A TEAM





A motivating fable: hiring a team





Defining the model

Definition (Stochastic Block Model)

Given a $q \times q$ matrix M of probabilities, associated SBM has

- q vertex classes of size n/q
- each edge (u, v) with independent probability M_{ij}, where i and j are u and v's respective classes.

We're interested in when all entries are c_{ij}/n for some constants c_{ij}.

Online Matching

Defining the model

Definition (Online Matching)

At each step: A random left vertex has all incident edges revealed, and the algorithm must choose which (if any) revealed edge to add to the matching. This choice cannot be undone.

Defining the model

Definition (Online Matching)

At each step: A random left vertex has all incident edges revealed, and the algorithm must choose which (if any) revealed edge to add to the matching. This choice cannot be undone.

Goal: find a strategy with a good ***expected competitive ratio*** – i.e. that maximizes the ratio of number of people you expect to match versus number of people a hirer with full info in advance could expect to match.

APPROACH 1: GREEDY

GREEDY: add a random available edge (if one exists)

Approach 1: GREEDY

GREEDY: add a random available edge (if one exists)

Definition (Equitable SBM)

A SBM is ***equitable*** if all classes have equal average degree.
That is, there's some constant c such that $\sum_j c_{ij} = c$ for all i.

Approach 1: GREEDY

GREEDY: add a random available edge (if one exists)

Definition (Equitable SBM)

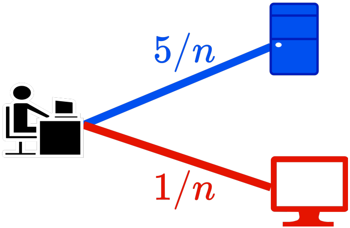
A SBM is ***equitable*** if all classes have equal average degree. That is, there's some constant c such that $\sum_j c_{ij} = c$ for all i.

Theorem

In equitable SBMs, GREEDY *achieves expected matching size* $\left(1 - \frac{\ln(2-e^{-c})}{c}\right) n$*, and this is optimal among online algorithms.*

*the case of bipartite Erdős–Rényi is a result of Mastin–Jaillet

Where GREEDY fails



APPROACH 2: DEGREEDY

DEGREEDY: match to the available class with smallest expected degree

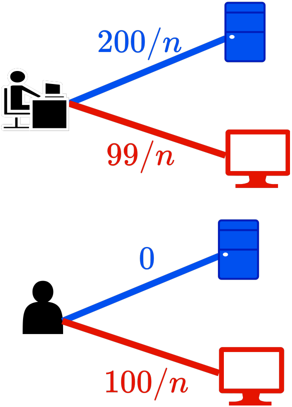
Approach 2: DEGREEDY

DEGREEDY: match to the available class with smallest expected degree

Proposition

When there is only one left class, DEGREEDY *is optimal among online algorithms.*

Where DEGREEDY fails



APPROACH 3: SHORTSIGHTED

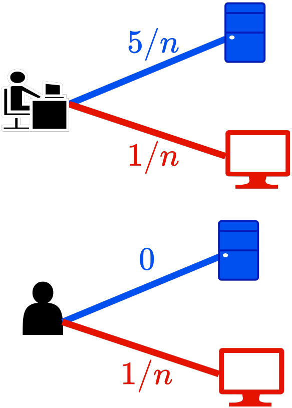
SHORTSIGHTED: match to the available class that maximizes probability of a match being available on the next step

APPROACH 3: SHORTSIGHTED

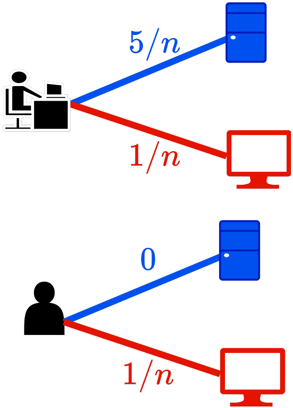
SHORTSIGHTED: match to the available class that maximizes probability of a match being available on the next step

Also optimal in certain cases (e.g. equitable), and performs well empirically on examples that we tried. However...

WHERE SHORTSIGHTED FAILS



WHERE SHORTSIGHTED FAILS



SHORTSIGHTED: $0.5749n$

Prefer blue jobs until 88% of vertices arrived: $0.5756n$

Approach 4: BRUTEFORCE

BRUTEFORCE: At the beginning, precompute optimal decision for every possible configuration (i.e. timestep, number of unmatched vertices in each class) by dynamic programming.

APPROACH 4: BRUTEFORCE

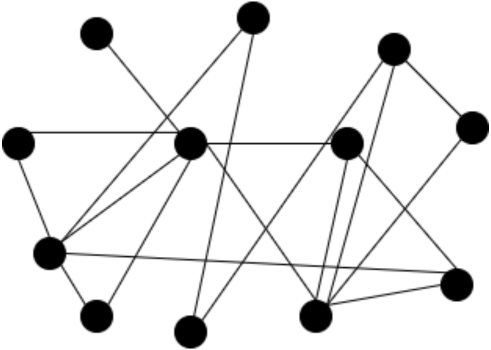
BRUTEFORCE: At the beginning, precompute optimal decision for every possible configuration (i.e. timestep, number of unmatched vertices in each class) by dynamic programming.

Optimal. But runtime $\Omega(n^{q+1})$, and appears difficult to analyze.

Offline Matching

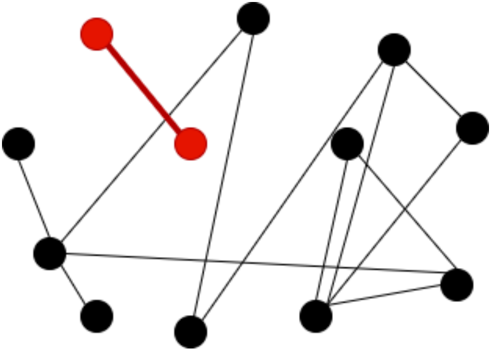
Karp–Sipser Algorithm

- If $\exists$ vertex of degree 1, add its edge to the matching.



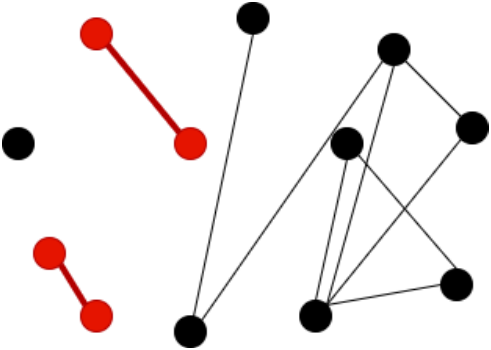
Karp–Sipser Algorithm

- If ∃ vertex of degree 1, choose one at random and add its edge to the matching.



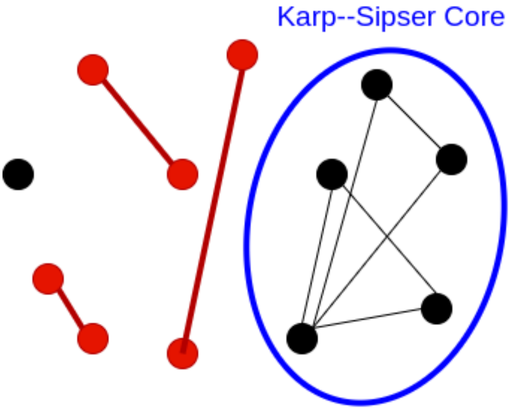
KARP–SIPSER ALGORITHM

- If ∃ vertex of degree 1, choose one at random and add its edge to the matching.



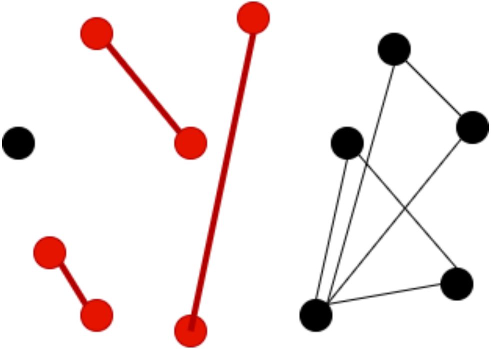
KARP–SIPSER ALGORITHM

- If ∃ vertex of degree 1, choose one at random and add its edge to the matching.



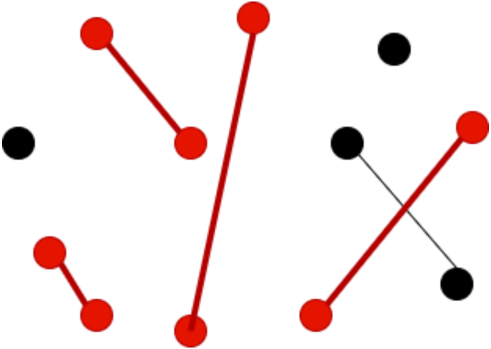
Karp–Sipser Algorithm

- If ∃ vertex of degree 1, choose one at random and add its edge to the matching.
- Otherwise, add any random edge



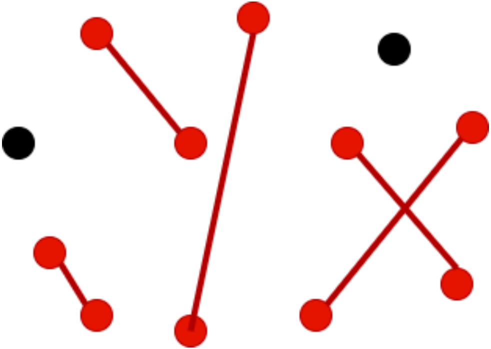
Karp–Sipser Algorithm

- If $\exists$ vertex of degree 1, choose one at random and add its edge to the matching.
- Otherwise, add any random edge



Karp–Sipser Algorithm

- If ∃ vertex of degree 1, choose one at random and add its edge to the matching.
- Otherwise, add any random edge



Karp–Sipser Behaviour in Erdős–Rényi Case

Optimality

Theorem (Karp, Sipser)

On an Erdős–Rényi graph with edge probability $\frac{c}{n}$, whp Karp–Sipser constructs a matching within $o(n)$ of optimal.

Phase Transition

Theorem (Karp, Sipser)

On an Erdős–Rényi graph with edge probability $\frac{c}{n}$, whp the Karp–Sipser core has size

- $o(n)$ *if* $c < e$
- $\Theta(n)$ *if* $c > e$

Comparisons to SBM Case

Optimality

- The Karp–Sipser is optimal in some specific cases, including the equitable case, which has the same matching number as Erdős–Rényi .
- It is **not** optimal on general SBM graphs

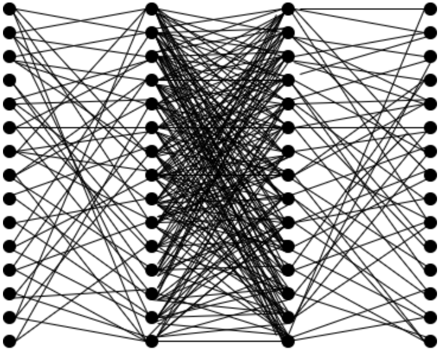
Phase Transition

- The Karp–Sipser core is $o(n)$ whenever average degree is $< e$ in all classes
- It is also $o(n)$ for more cases

Where Karp–Sipser Fails

$$c_{ab} = c_{cd} = 10$$

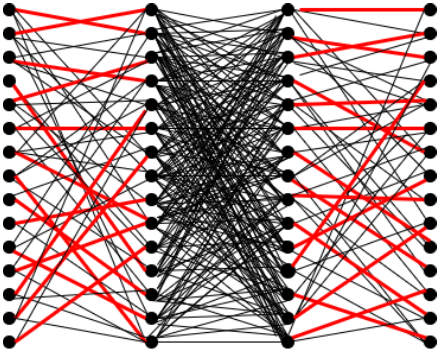
$$c_{bc} = 100$$



Where Karp–Sipser Fails

$$c_{ab} = c_{cd} = 10$$

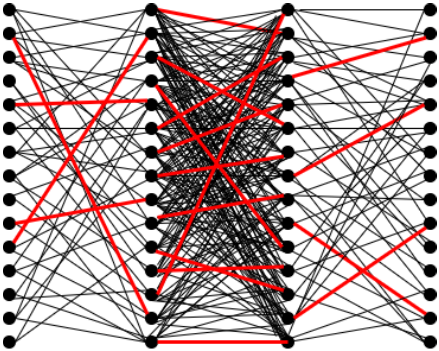
$$c_{bc} = 100$$



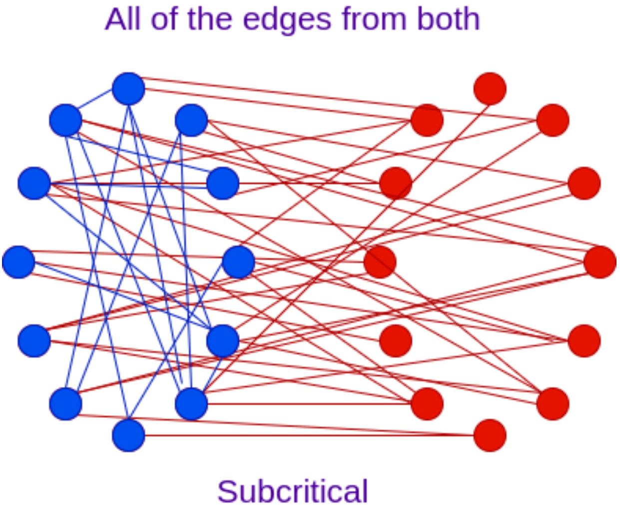
Where Karp–Sipser Fails

$$c_{ab} = c_{cd} = 10$$

$$c_{bc} = 100$$



Critical Threshold



QUESTIONS

- Is there a description of the critical threshold?
- Is a label-aware version of Karp–Sipser optimal?
- More generally, can we describe a procedure that, given the probability matrix of a SBM, determines the asysmptotic matching number?
- In the online setting, does SHORTSIGHTED always achieve competitive ratio close to BRUTE-FORCE?
- Does there exist a linear-time online algorithm with the same competitive ratio as BRUTE-FORCE?

Thank you!