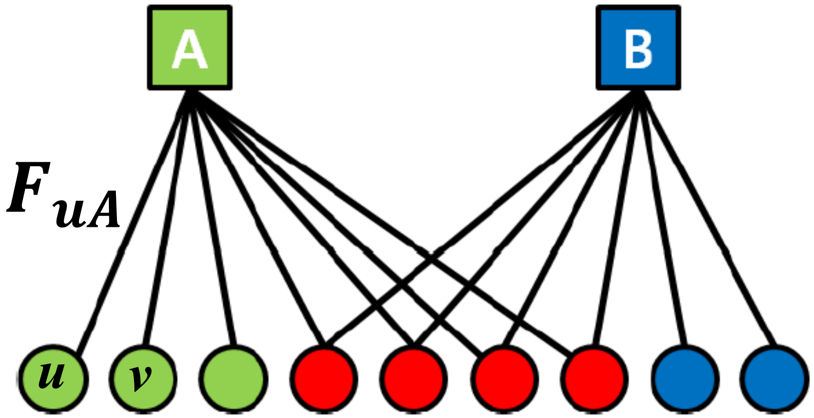
From AGM to BIGCLAM

Mining of Massive Datasets
Leskovec, Rajaraman, and Ullman
Stanford University

From AGM to BigCLAM

- **Relaxation: Memberships have strengths**



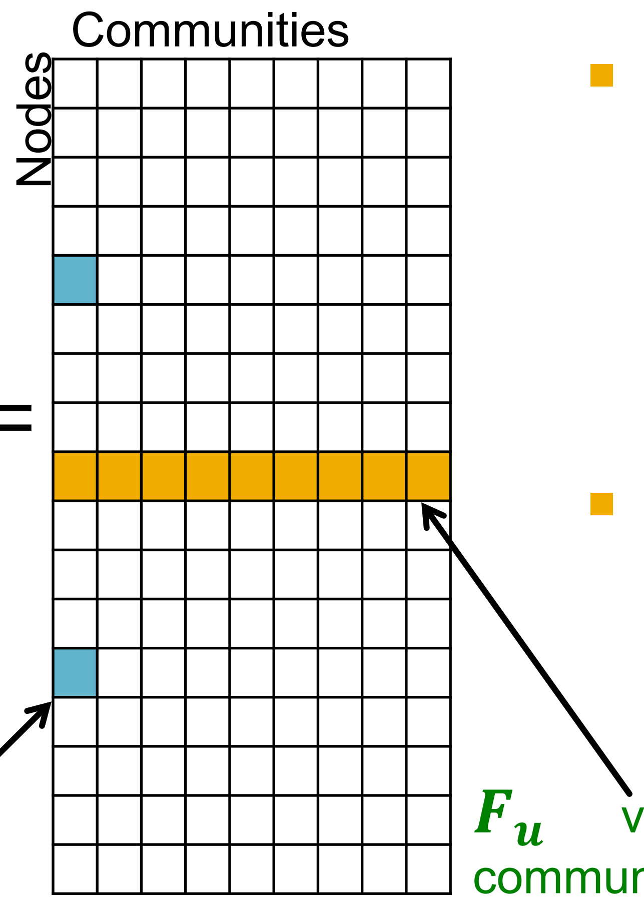
- F_{uA}: The membership strength of node u to community A ($F_{uA} = 0$: no membership)
- **Each community A links nodes independently:**

$$P_A(u,v) = 1 - \exp(-F_{uA} \cdot F_{vA})$$

Factor Matrix F

- **Community membership strength matrix F**

$F =$



F_{vA}
strength of u's membership to A

F_u vector of community membership strengths of u

- $P_A(u,v) = 1 - \exp(-F_{uA} \cdot F_{vA})$
  - **Probability of connection is proportional to the product of strengths**
    - **Notice:** If one node doesn't belong to the community ($F_{uC} = 0$) then $P(u,v) = 0$
- Prob. that **at least one** common community C links the nodes:
  - $P(u,v) = 1 - \prod_C (1 - P_C(u,v))$

From AGM to BigCLAM

- Community $\boldsymbol{A}$ links nodes $\boldsymbol{u}, \boldsymbol{v}$ independently:

$$P_A(u,v) = 1 - \exp(-F_{uA} \cdot F_{vA})$$

- Then prob. at least one common $\boldsymbol{C}$ links them:

$$P(u,v) = 1 - \prod_C \big(1 - P_C(u,v)\big)$$
$$= 1 - \exp(-\textstyle\sum_C F_{uC} \cdot F_{vC})$$
$$= 1 - \exp(-F_u \cdot F_v^T)$$

- **For example:**

F_u:	0	1.2	0	0.2
F_v:	0.5	0	0	0.8
F_w:	0	1.8	1	0

Node community membership strengths

Then: $F_u \cdot F_v^T = 0.16$

And: $P(u,v) = 1 - exp(-0.16) = 0.14$

But: $P(u,w) = 0.88$

$P(v,w) = 0$