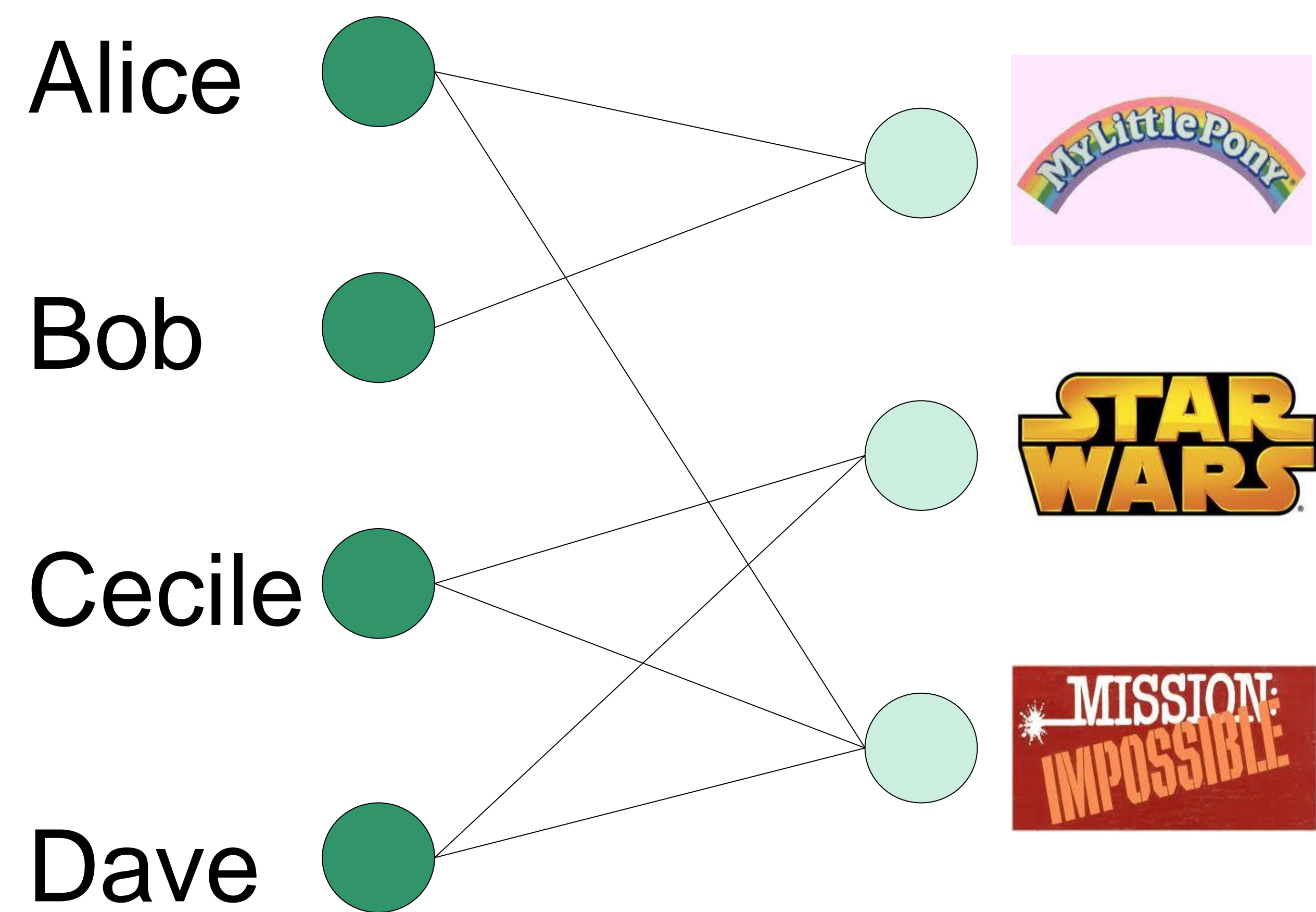


Reconstructing Graphs from Neighborhood Data

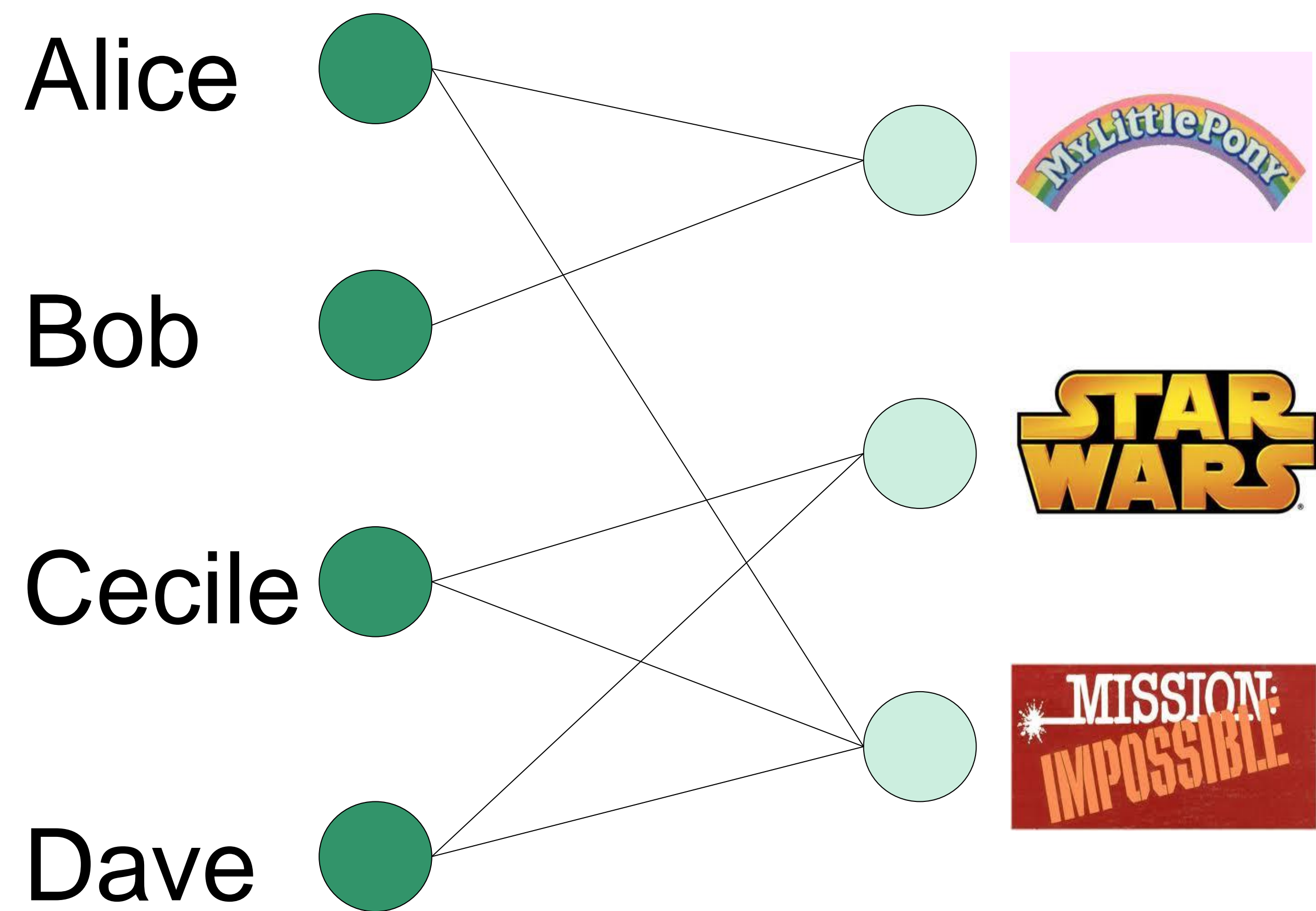
Dora Erdos, Rainer Gemulla, Evimaria Terzi





			
Alice	1	0	1
Bob	1	0	0
Cecile	0	1	1
Dave	0	1	1

M

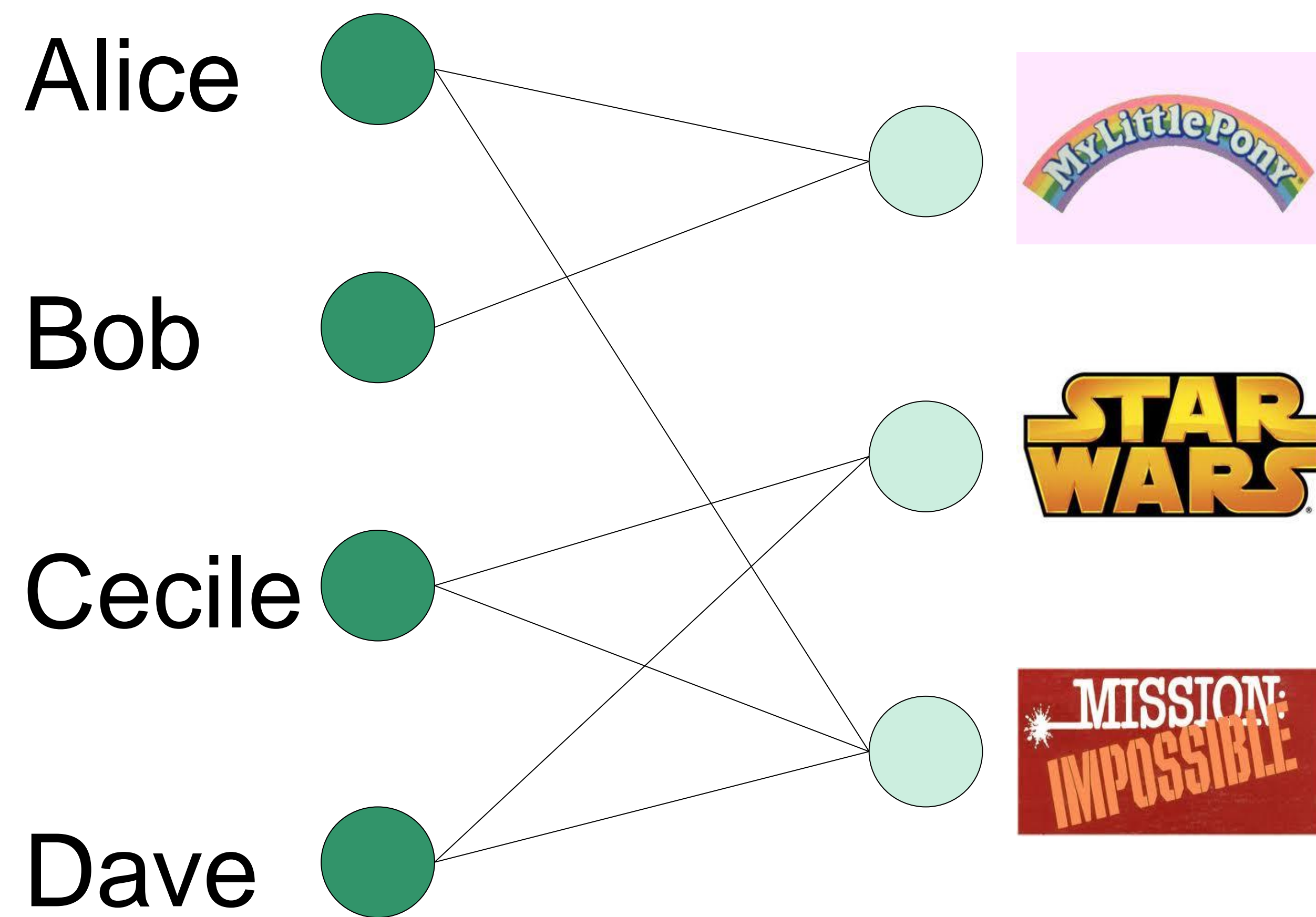


	Alice	Bob	Cecile	Dave
Alice	2	1	1	1
Bob	1	1	0	0
Cecile	1	0	2	2
Dave	1	0	2	2

L

			
Alice	1	0	1
Bob	1	0	0
Cecile	0	1	1
Dave	0	1	1

M



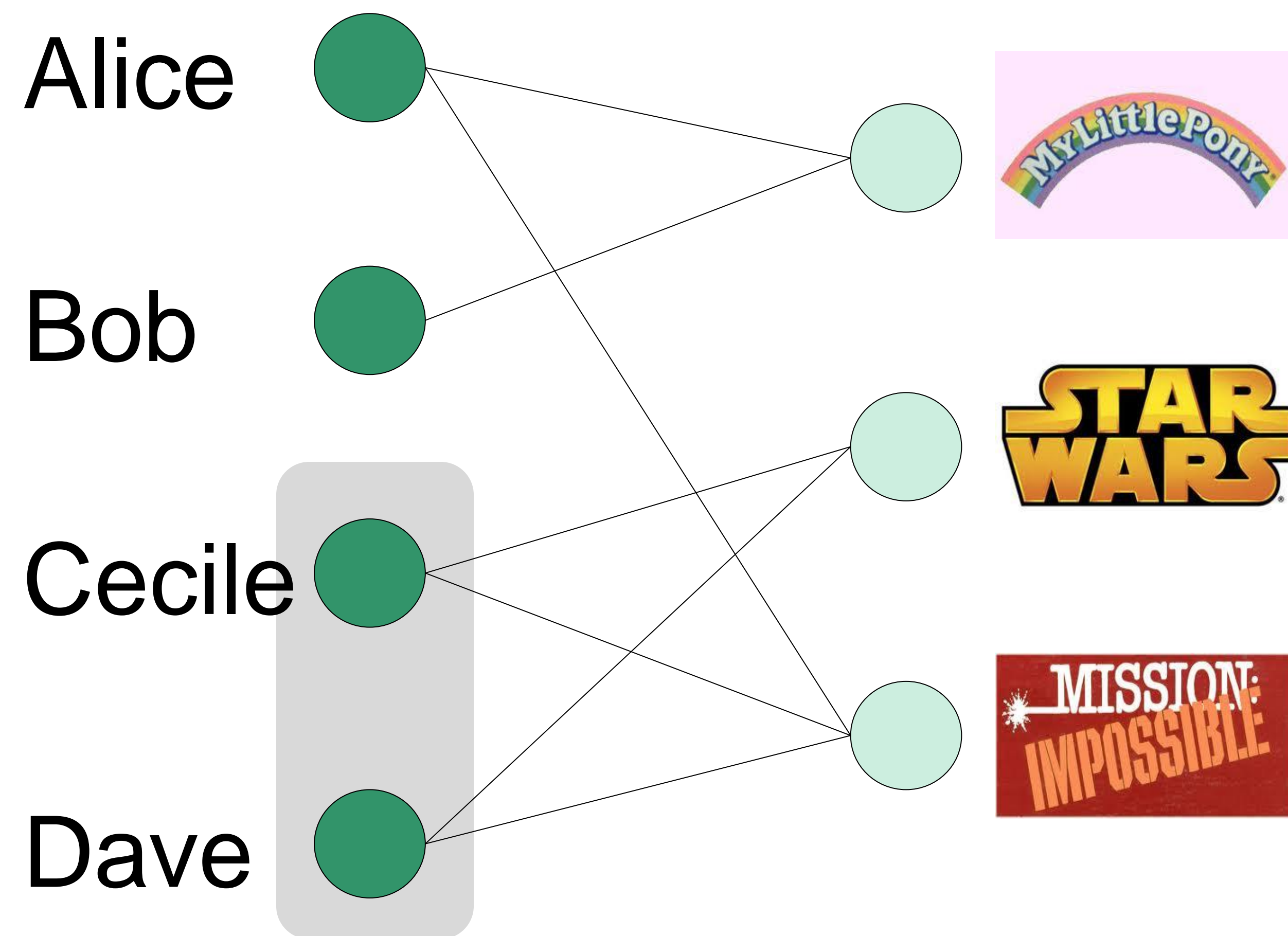
Left neighborhood information of M.

	Alice	Bob	Cecile	Dave
Alice	2	1	1	1
Bob	1	1	0	0
Cecile	1	0	2	2
Dave	1	0	2	2

L

			
Alice	1	0	1
Bob	1	0	0
Cecile	0	1	1
Dave	0	1	1

M



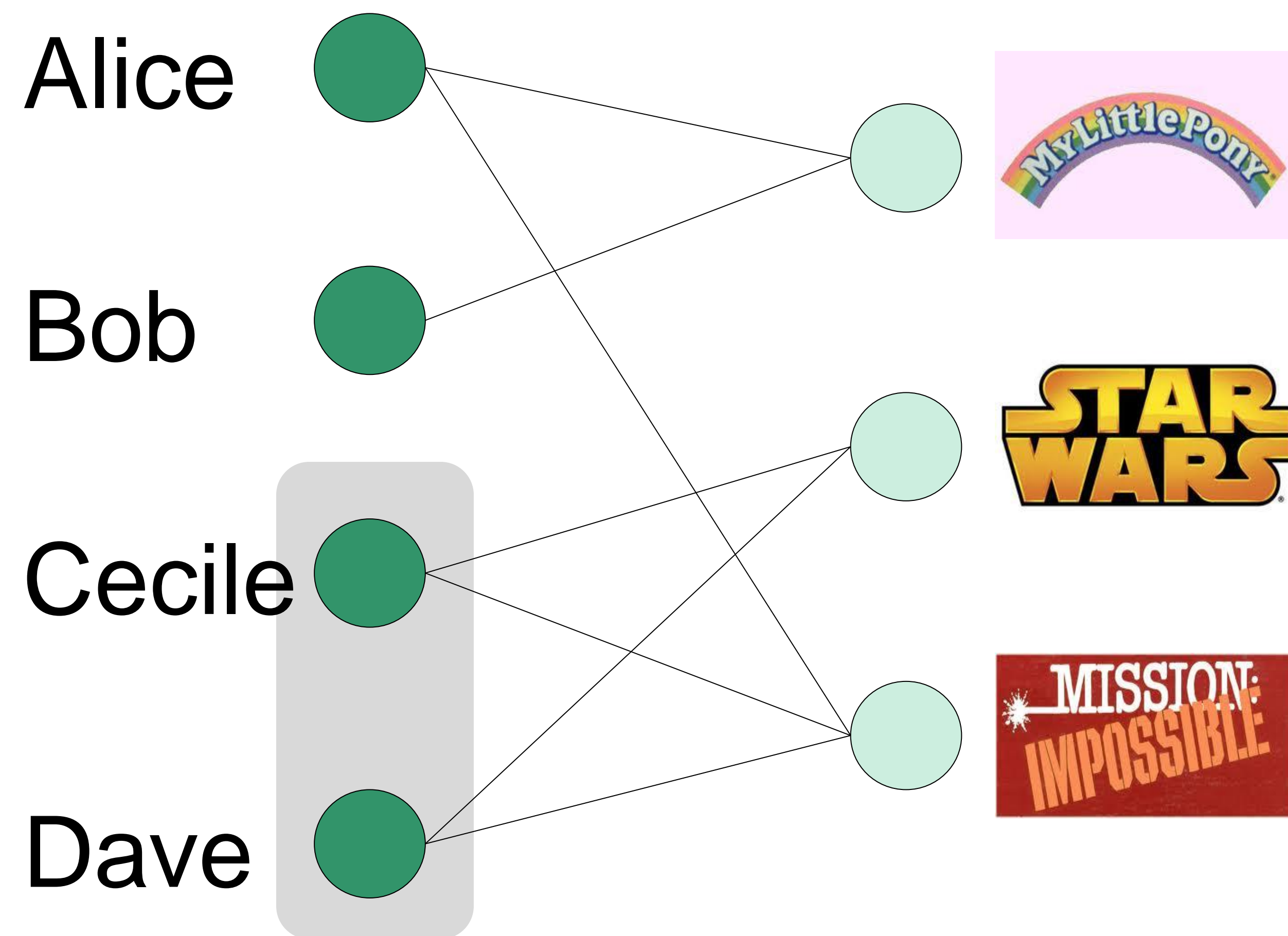
Left neighborhood information of M.

	Alice	Bob	Cecile	Dave
Alice	2	1	1	1
Bob	1	1	0	0
Cecile	1	0	2	2
Dave	1	0	2	2

L

			
Alice	1	0	1
Bob	1	0	0
Cecile	0	1	1
Dave	0	1	1

M



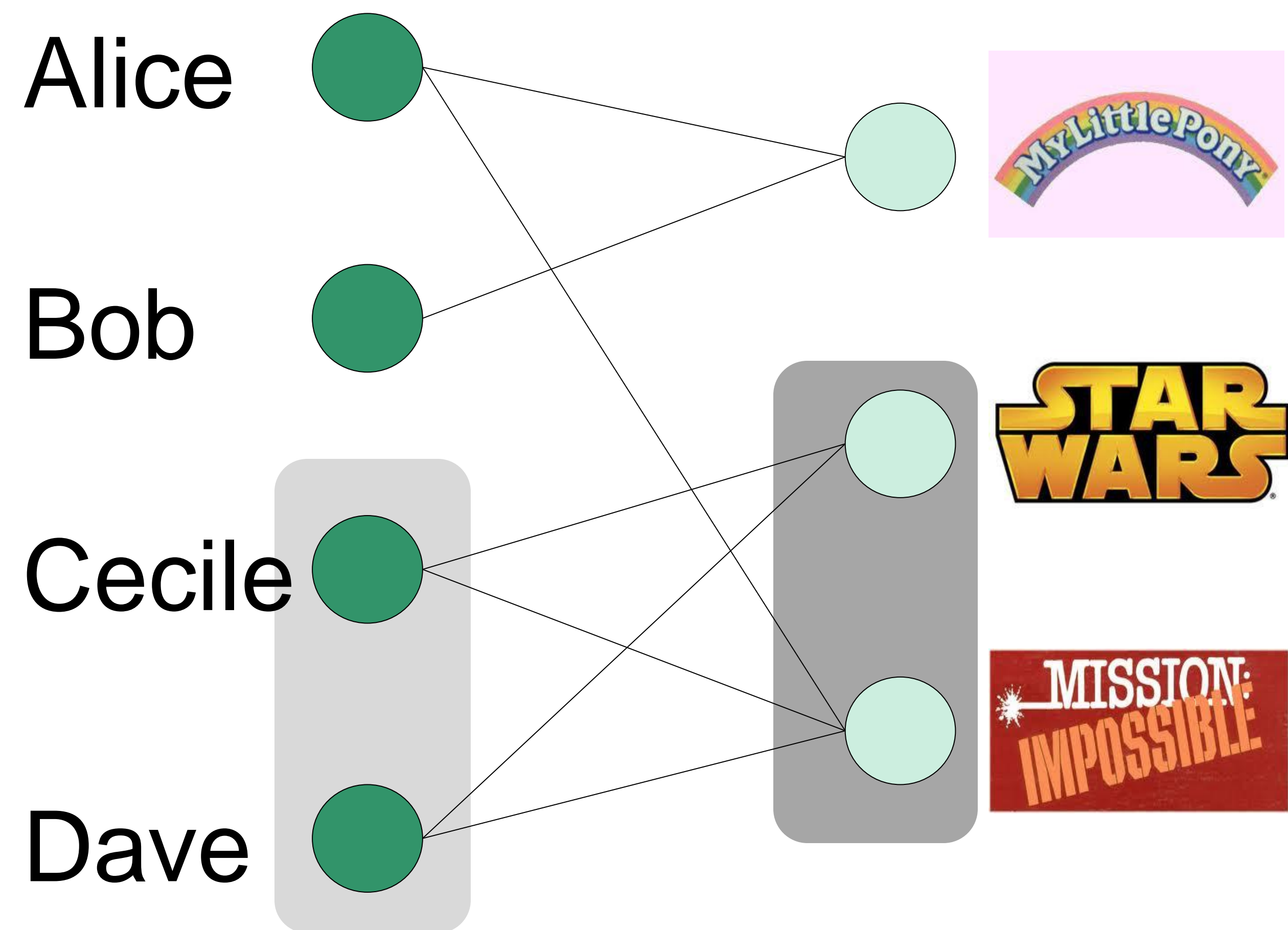
Left neighborhood information of M.

	Alice	Bob	Cecile	Dave
Alice	2	1	1	1
Bob	1	1	0	0
Cecile	1	0	2	2
Dave	1	0	2	2

L

			
Alice	1	0	1
Bob	1	0	0
Cecile	0	1	1
Dave	0	1	1

M



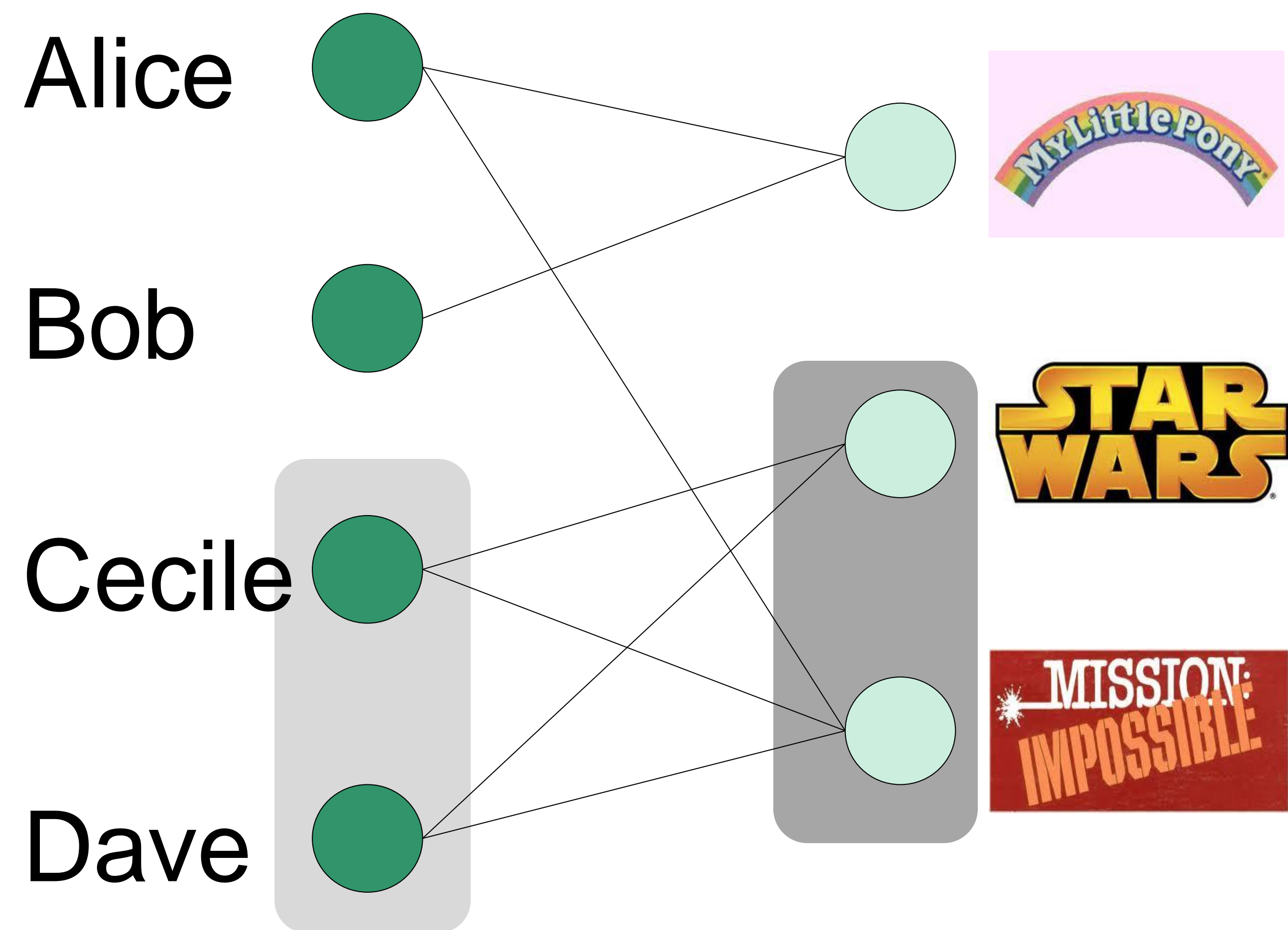
Left neighborhood information of M.

	Alice	Bob	Cecile	Dave
Alice	2	1	1	1
Bob	1	1	0	0
Cecile	1	0	2	2
Dave	1	0	2	2

L

			
Alice	1	0	1
Bob	1	0	0
Cecile	0	1	1
Dave	0	1	1

M



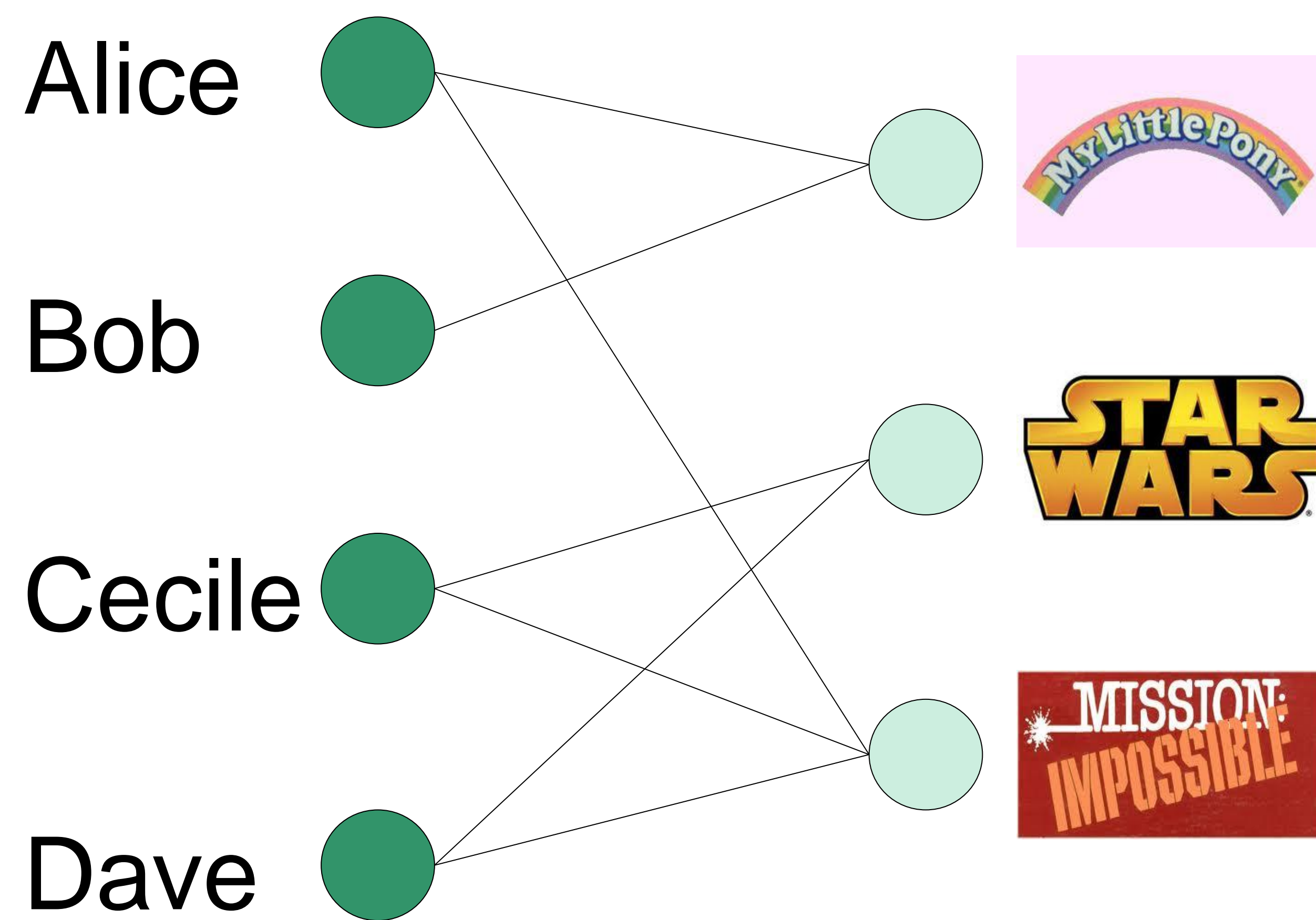
Left neighborhood information of M.

	Alice	Bob	Cecile	Dave
Alice	2	1	1	1
Bob	1	1	0	0
Cecile	1	0	2	2
Dave	1	0	2	2

L

			
Alice	1	0	1
Bob	1	0	0
Cecile	0	1	1
Dave	0	1	1

M



Left neighborhood information of M.

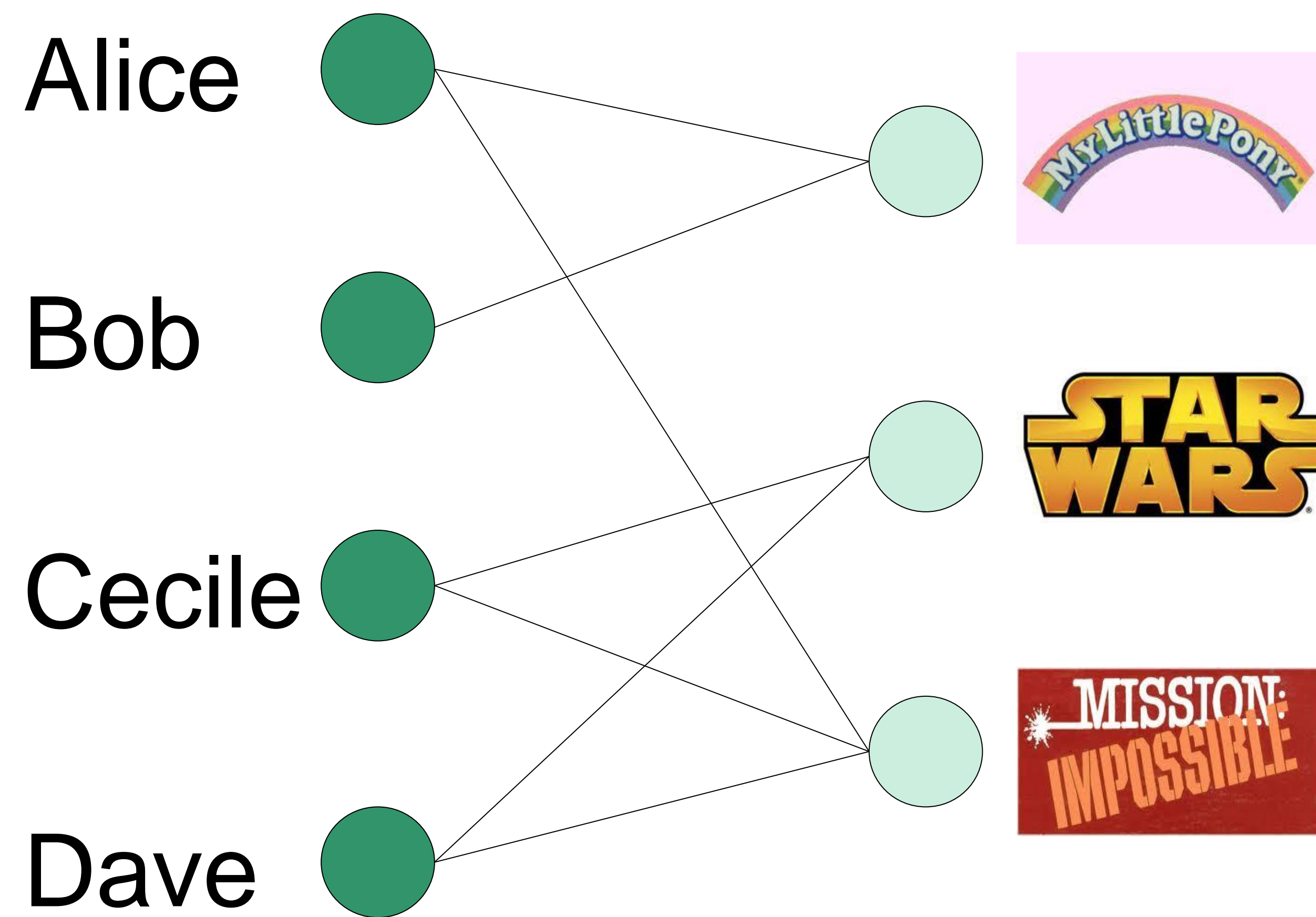
L is a similarity matrix between people

	Alice	Bob	Cecile	Dave
Alice	2	1	1	1
Bob	1	1	0	0
Cecile	1	0	2	2
Dave	1	0	2	2

L

			
Alice	1	0	1
Bob	1	0	0
Cecile	0	1	1
Dave	0	1	1

M



Right neighborhood information of M.

			
	2	0	1
	0	2	2
	1	2	3

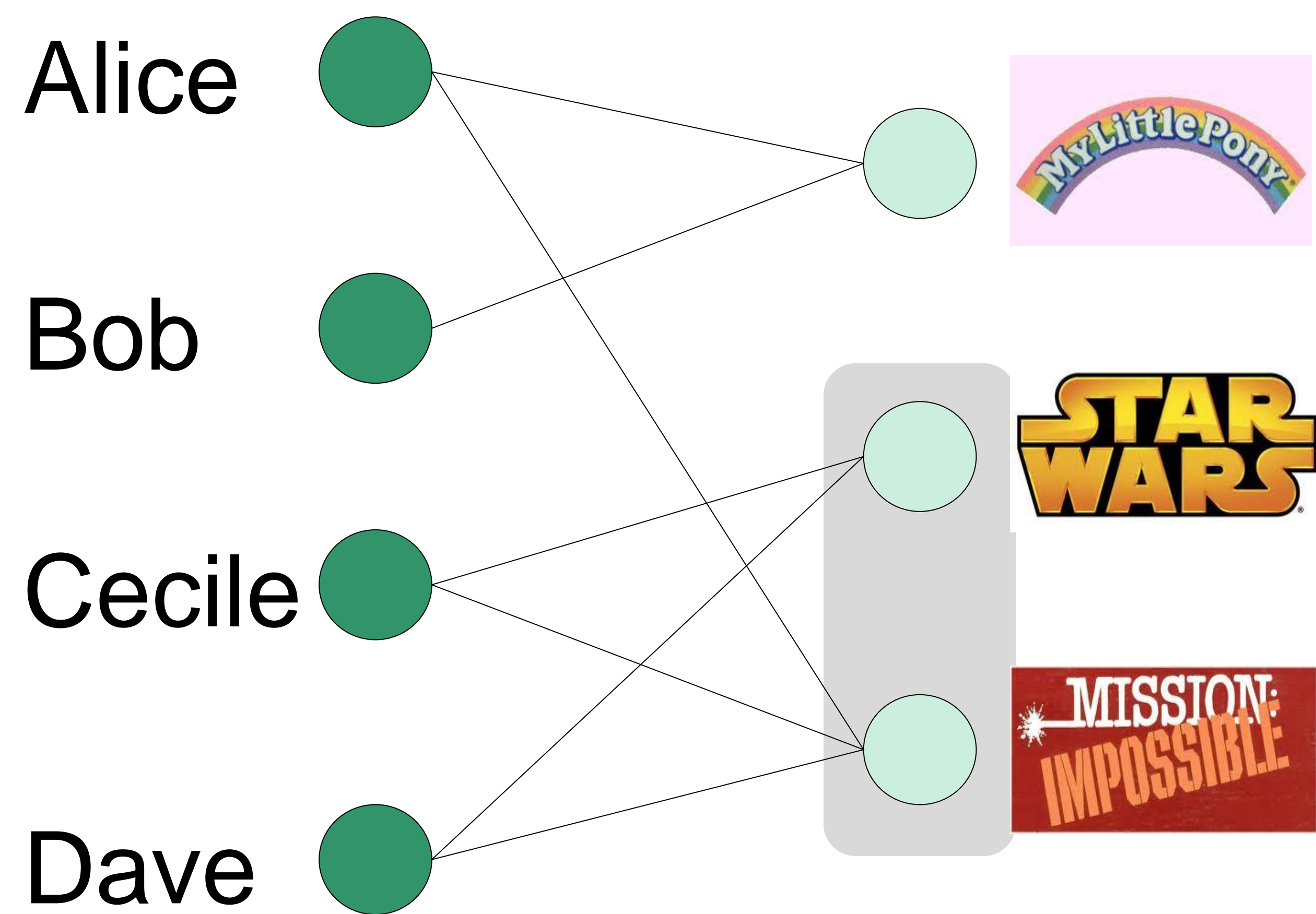
R

	Alice	Bob	Cecile	Dave
Alice	2	1	1	1
Bob	1	1	0	0
Cecile	1	0	2	2
Dave	1	0	2	2

L

			
Alice	1	0	1
Bob	1	0	0
Cecile	0	1	1
Dave	0	1	1

M



Right neighborhood information of M.

			
	2	0	1
	0	2	2
	1	2	3

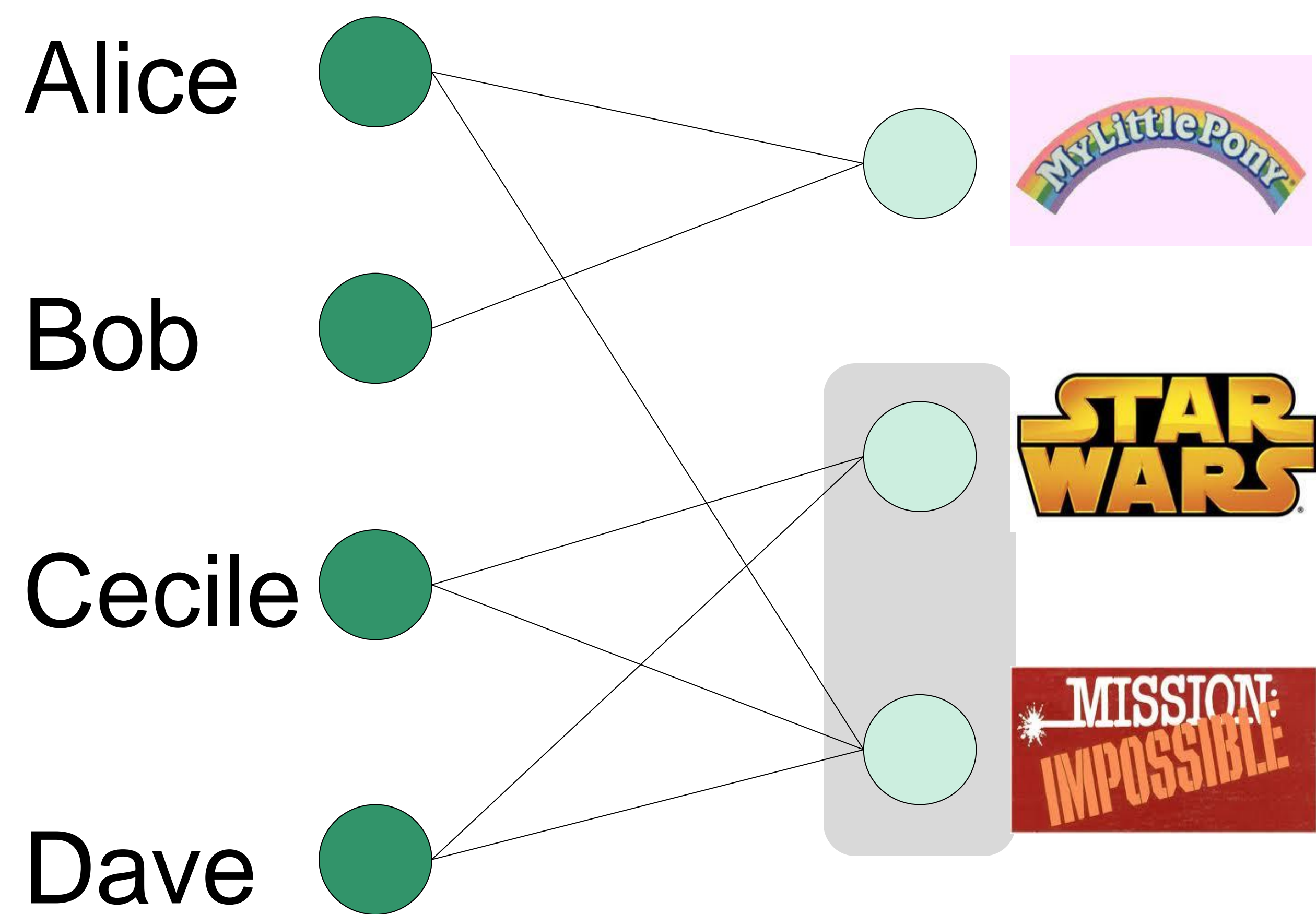
R

	Alice	Bob	Cecile	Dave
Alice	2	1	1	1
Bob	1	1	0	0
Cecile	1	0	2	2
Dave	1	0	2	2

L

			
Alice	1	0	1
Bob	1	0	0
Cecile	0	1	1
Dave	0	1	1

M



Right neighborhood information of M.

			
	2	0	1
	0	2	2
	1	2	3

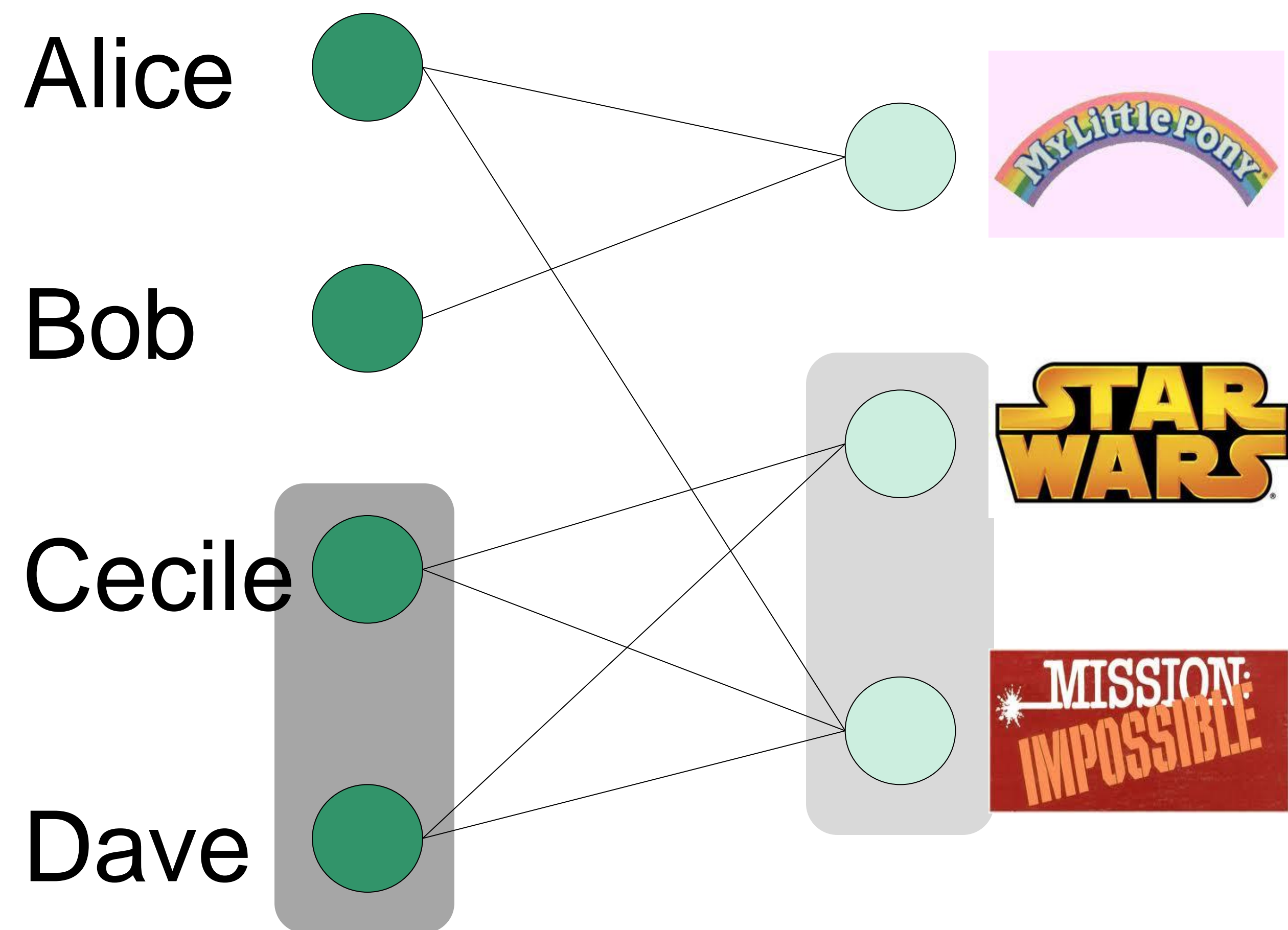
R

	Alice	Bob	Cecile	Dave
Alice	2	1	1	1
Bob	1	1	0	0
Cecile	1	0	2	2
Dave	1	0	2	2

L

			
Alice	1	0	1
Bob	1	0	0
Cecile	0	1	1
Dave	0	1	1

M



Right neighborhood information of M.

			
	2	0	1
	0	2	2
	1	2	3

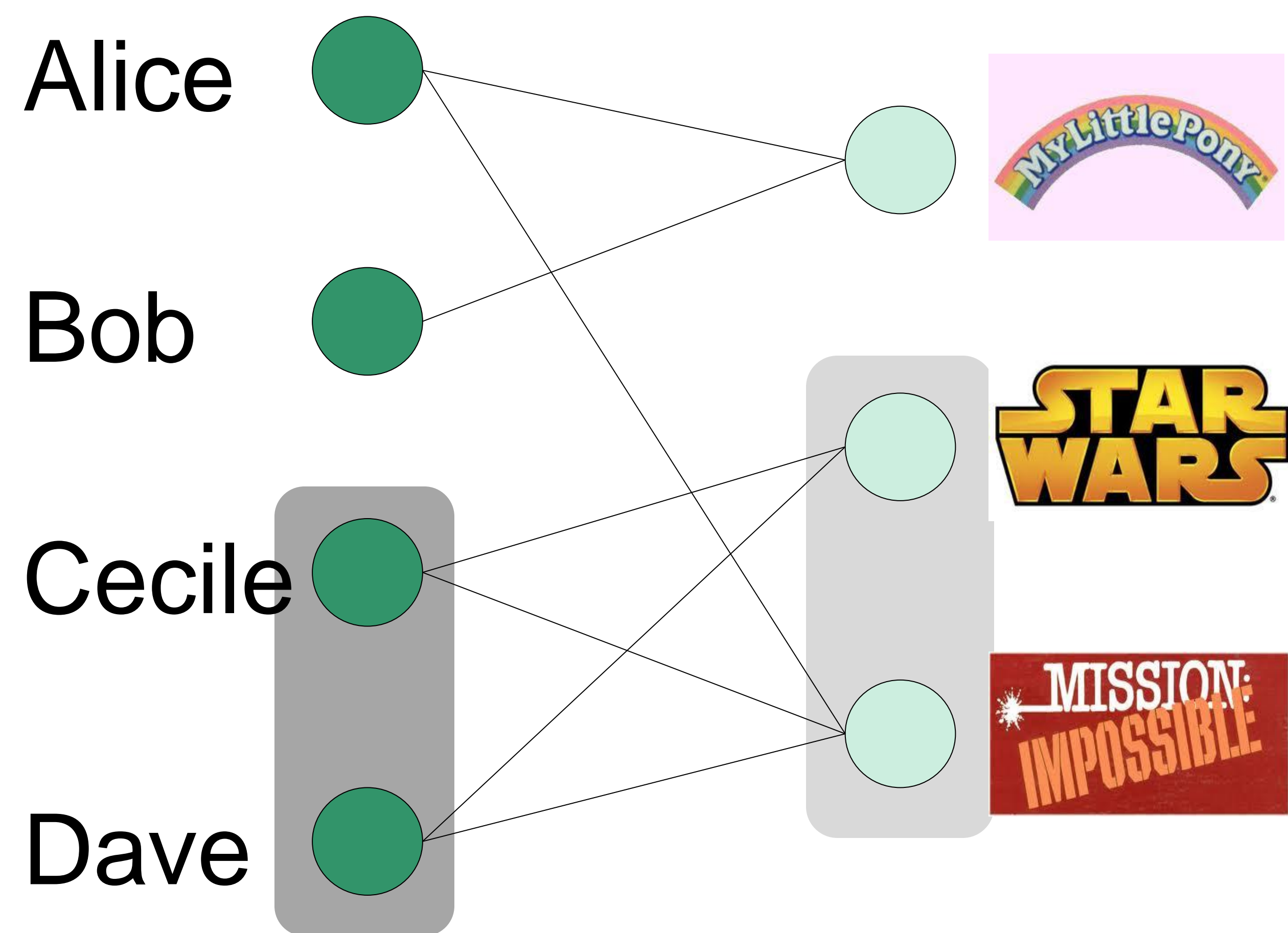
R

	Alice	Bob	Cecile	Dave
Alice	2	1	1	1
Bob	1	1	0	0
Cecile	1	0	2	2
Dave	1	0	2	2


L

			
Alice	1	0	1
Bob	1	0	0
Cecile	0	1	1
Dave	0	1	1

M



Right neighborhood information of M.

			
	2	0	1
	0	2	2
	1	2	3

R

	Alice	Bob	Cecile	Dave
Alice	2	1	1	1
Bob	1	1	0	0
Cecile	1	0	2	2
Dave	1	0	2	2

L

L is a similarity matrix between people
 R is a similarity matrix between movies

			
Alice	1	0	1
Bob	1	0	0
Cecile	0	1	1
Dave	0	1	1

M

			
	2	0	1
	0	2	2
	1	2	3

R

	Alice	Bob	Cecile	Dave
Alice	2	1	1	1
Bob	1	1	0	0
Cecile	1	0	2	2
Dave	1	0	2	2

L

L is a similarity matrix between people

R is a similarity matrix between movies



Recommend movies to users

			
Alice	1	0	1
Bob	1	0	0
Cecile	0	1	1
Dave	0	1	1

M

			
	2	0	1
	0	2	2
	1	2	3

R

	Alice	Bob	Cecile	Dave
Alice	2	1	1	1
Bob	1	1	0	0
Cecile	1	0	2	2
Dave	1	0	2	2

L

L is a similarity matrix between people

R is a similarity matrix between movies



Recommend movies to users

Recommendation systems:

1. Compute similarity matrices L and R from M

$$L = MM^T$$

$$R = M^T M$$

2. Apply some algorithm to L and R to obtain recommendation

collaborative filtering

nearest neighbor algorithm

			
	2	0	1
	0	2	2
	1	2	3

R

	Alice	Bob	Cecile	Dave
Alice	2	1	1	1
Bob	1	1	0	0
Cecile	1	0	2	2
Dave	1	0	2	2

L

Recommendation systems:

1. Compute similarity matrices L and R from M

$$L = MM^T$$

$$R = M^T M$$

2. Apply some algorithm to L and R to obtain recommendation

collaborative filtering
nearest neighbor algorithm

What do we do if M is hidden?

			
	2	0	1
	0	2	2
	1	2	3

R

	Alice	Bob	Cecile	Dave
Alice	2	1	1	1
Bob	1	1	0	0
Cecile	1	0	2	2
Dave	1	0	2	2

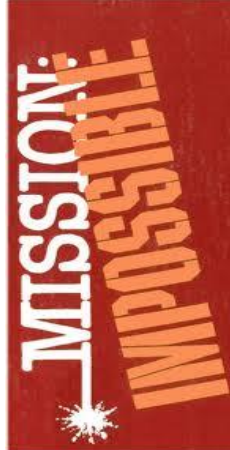

L



What do we do if M is hidden?

Reconstruction problem:

Given L and R reconstruct M.

			
	2	0	1
	0	2	2
	1	2	3

R

Outline

Problem definition

Connection to SVD

Greedy SVD reconstruction

Experiments

Discussion

Outline

Problem definition

Connection to SVD

Greedy SVD reconstruction

Experiments

Discussion

	Alice	Bob	Cecile	Dave
Alice	2	1	1	1
Bob	1	1	0	0
Cecile	1	0	2	2
Dave	1	0	2	2

L

			
Alice	1	0	1
Bob	1	0	0
Cecile	0	1	1
Dave	0	1	1

M


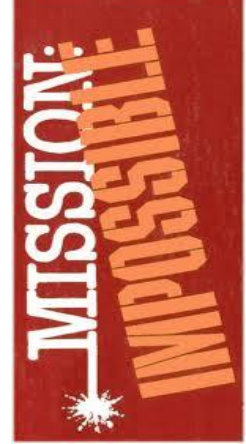
Given the neighborhood information L and R how would you reconstruct M?

			
	2	0	1
	0	2	2
	1	2	3

R

	Alice	Bob	Cecile	Dave
Alice	2	1	1	1
Bob	1	1	0	0
Cecile	1	0	2	2
Dave	1	0	2	2

L



			
Alice	1	0	1
Bob	1	0	0
Cecile	0	1	1
Dave	0	1	1

M

Given the neighborhood information L and R how would you reconstruct M?

$$L = MM^T$$

$$R = M^T M$$

			
	2	0	1
	0	2	2
	1	2	3

R

Reconstruction problem:

Given neighborhood matrices L and R construct binary matrix \hat{M}

Such that $F_L(\hat{M}) + F_R(\hat{M})$ is minimized

$$F_L(\hat{M}) = \left\| \hat{M}\hat{M}^T - L \right\|$$

$$F_R(\hat{M}) = \left\| \hat{M}^T \hat{M} - R \right\|$$

Measures the distance between L , R and the neighborhood matrices of \hat{M}

Outline

Problem definition

Connection to SVD

Greedy SVD reconstruction

Experiments

Discussion

Connection to SVD

Idea: Reconstruct M by obtaining its SVD decomposition

Connection to SVD

Idea: Reconstruct M by obtaining its SVD decomposition

SVD decomposition of M : $M = U\Sigma V^T$

Connection to SVD

Idea: Reconstruct M by obtaining its SVD decomposition

SVD decomposition of M : $M = U\Sigma V^T$

Eigen decomposition of L : $L = U\Lambda U^T$ Remember: $L = MM^T$

Connection to SVD

Idea: Reconstruct M by obtaining its SVD decomposition

SVD decomposition of M : $M = U\Sigma V^T$

Eigen decomposition of L : $L = U\Lambda U^T$ Remember: $L = MM^T$

Eigen decomposition of R : $R = V\Lambda V^T$ $R = M^T M$

Connection to SVD

Idea: Reconstruct M by obtaining its SVD decomposition

SVD decomposition of M : $M = U\Sigma V^T$

Eigen decomposition of L : $L = U\Lambda U^T$ Remember: $L = MM^T$

Eigen decomposition of R : $R = V\Lambda V^T$ $R = M^T M$

Observe $\Lambda = \Sigma^2$

Connection to SVD

Eigen decomposition of L and R: $L = U\Lambda U^T$ $R = V\Lambda V^T$ $\Lambda = \Sigma^2$

As a result we have all components of the SVD representation of M

$$U, V \text{ and } \Sigma = \sqrt{\Lambda}$$

Connection to SVD

Eigen decomposition of L and R: $L = U\Lambda U^T$ $R = V\Lambda V^T$ $\Lambda = \Sigma^2$

As a result we have all components of the SVD representation of M

$$U, V \text{ and } \Sigma = \sqrt{\Lambda}$$

Compute $\hat{M} = U\sqrt{\Lambda}V^T$

Connection to SVD

Eigen decomposition of L and R: $L = U\Lambda U^T$ $R = V\Lambda V^T$ $\Lambda = \Sigma^2$

As a result we have all components of the SVD representation of M

$$U, V \text{ and } \Sigma = \sqrt{\Lambda}$$

Compute $\hat{M} = U\sqrt{\Lambda}V^T$

Done?

Connection to SVD

Eigen decomposition of L and R: $L = U\Lambda U^T$ $R = V\Lambda V^T$ $\Lambda = \Sigma^2$

As a result we have all components of the SVD representation of M

$$U, V \text{ and } \Sigma = \sqrt{\Lambda}$$

Compute $\hat{M} = U\sqrt{\Lambda}V^T$
Done?

We don't know what sign to pick for the singular values

$$\hat{\Sigma}_i = \pm\sqrt{\Lambda}_i$$

We don't know whether the resulting \hat{M} is binary.

Outline

Problem definition

Connection to SVD

Greedy SVD reconstruction

Experiments

Discussion

Greedy SVD reconstruction

1. Compute eigenvalue decompositions of L and R to obtain U, V, Λ

$$\text{Set } \hat{\Sigma}_i = \pm\sqrt{\Lambda_i}$$

2. Fix signs of singular values in decreasing order of magnitude

$$|\Sigma_1| \geq |\Sigma_2| \geq \dots |\Sigma_n|$$

Greedy SVD reconstruction

1. Compute eigenvalue decompositions of L and R to obtain U, V, Λ

$$\text{Set } \hat{\Sigma}_i = \pm\sqrt{\Lambda_i}$$

2. Fix signs of singular values in decreasing order of magnitude

$$|\Sigma_1| \geq |\Sigma_2| \geq \dots \geq |\Sigma_n|$$

Iteration i :

$$M_i^+ = M_{i-1} + U_i \cdot |\Sigma_i| \cdot V_i^T$$

$$M_i^- = M_{i-1} - U_i \cdot |\Sigma_i| \cdot V_i^T$$

Greedy SVD reconstruction

1. Compute eigenvalue decompositions of L and R to obtain U, V, Λ

$$\text{Set } \hat{\Sigma}_i = \pm\sqrt{\Lambda}$$

2. Fix signs of singular values in decreasing order of magnitude

$$|\Sigma_1| \geq |\Sigma_2| \geq \dots |\Sigma_n|$$

Iteration i :

$$M_i^+ = M_{i-1} + U_i \cdot |\Sigma_i| \cdot V_i^T$$

$$M_i^- = M_{i-1} - U_i \cdot |\Sigma_i| \cdot V_i^T$$

Compute binary matrix by rounding:

$$BM_i^+ = (M_i^+ \geq t)$$

$$BM_i^- = (M_i^- \geq t)$$

Greedy SVD reconstruction

1. Compute eigenvalue decompositions of L and R to obtain U, V, Λ

$$\text{Set } \hat{\Sigma}_i = \pm \sqrt{\Lambda_i}$$

2. Fix signs of singular values in decreasing order of magnitude

$$|\Sigma_1| \geq |\Sigma_2| \geq \dots \geq |\Sigma_n|$$

Iteration i :

$$M_i^+ = M_{i-1} + U_i \cdot |\Sigma_i| \cdot V_i^T$$

Compute binary matrix by rounding:

$$M_i^- = M_{i-1} - U_i \cdot |\Sigma_i| \cdot V_i^T$$

$$BM_i^+ = (M_i^+ \geq t)$$

Choose binary matrix that is closest to its real version

$$BM_i^- = (M_i^- \geq t)$$

Greedy SVD reconstruction

$t = 0.5$ Closest binary matrix

$t = 0.1$ Predicts mostly 1s

$t = 0.9$ Predicts mostly 0s

Iteration i :

$$M_i^+ = M_{i-1} + U_i \cdot |\Sigma_i| \cdot V_i^T$$

Compute binary matrix by rounding:

$$M_i^- = M_{i-1} - U_i \cdot |\Sigma_i| \cdot V_i^T$$

$$BM_i^+ = (M_i^+ \geq t)$$

Choose binary matrix that is closest to its real version

$$BM_i^- = (M_i^- \geq t)$$

Outline

Problem definition

Connection to SVD

Greedy SVD reconstruction

Experiments

Discussion

Experiments

Flickr dataset¹:

2000 users x 1989 groups

Density 5%

Users exhibit power law degree distribution

Groups have exponential degree distribution

¹Zheleva *et al.* WWW '09

Experiments

Flickr dataset¹: 2000 users x 1989 groups
Density 5%

Data is very sparse

estimating M to be all 0 would yield very
small error

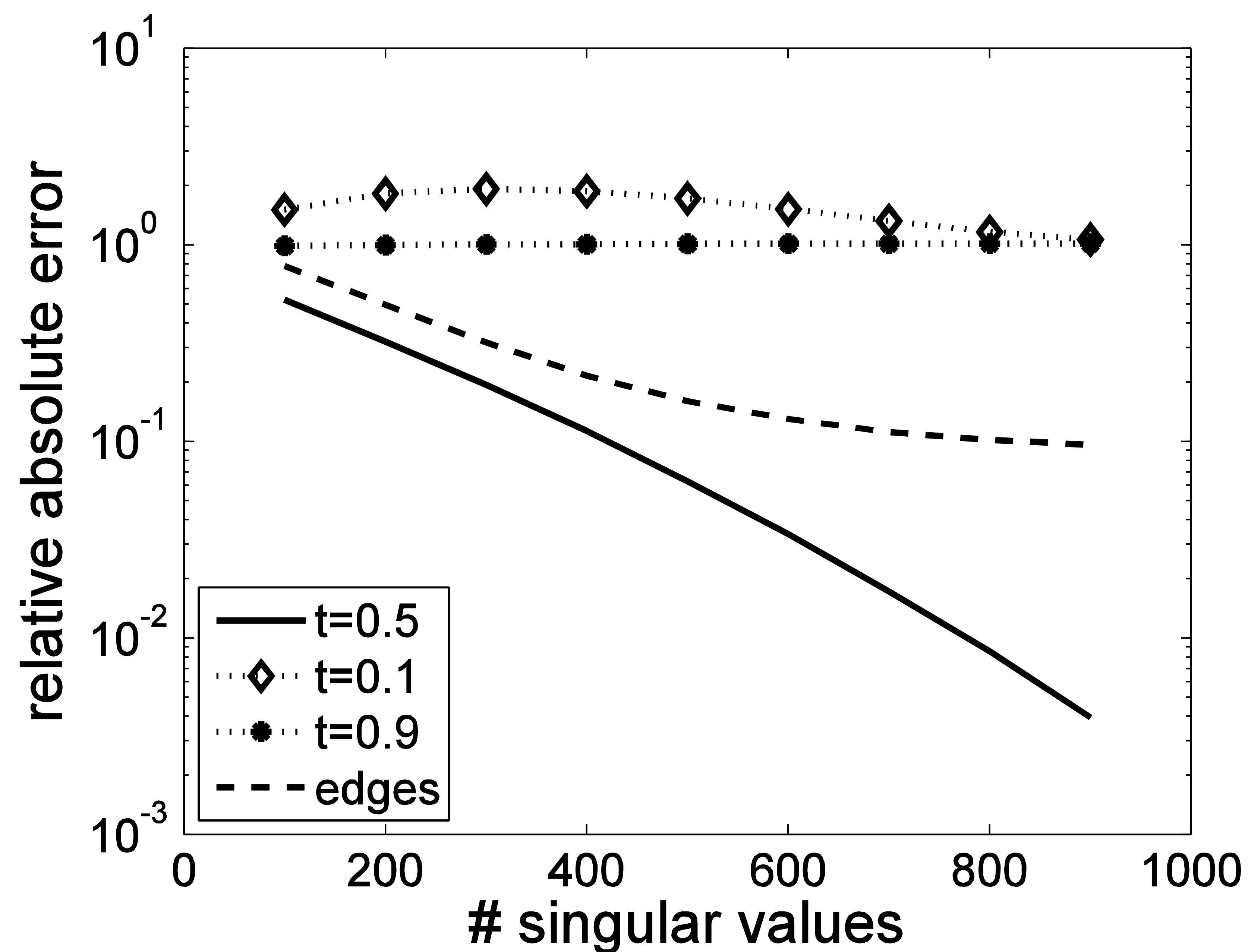
Relative absolute error

$$\frac{\|B\hat{M} - M\|}{\|M\|}$$

¹Zheleva *et al.* WWW '09

Experiments

Flickr dataset¹: 2000 users x 1989 groups
Density 5%



X axis: number k of highest magnitude singular values

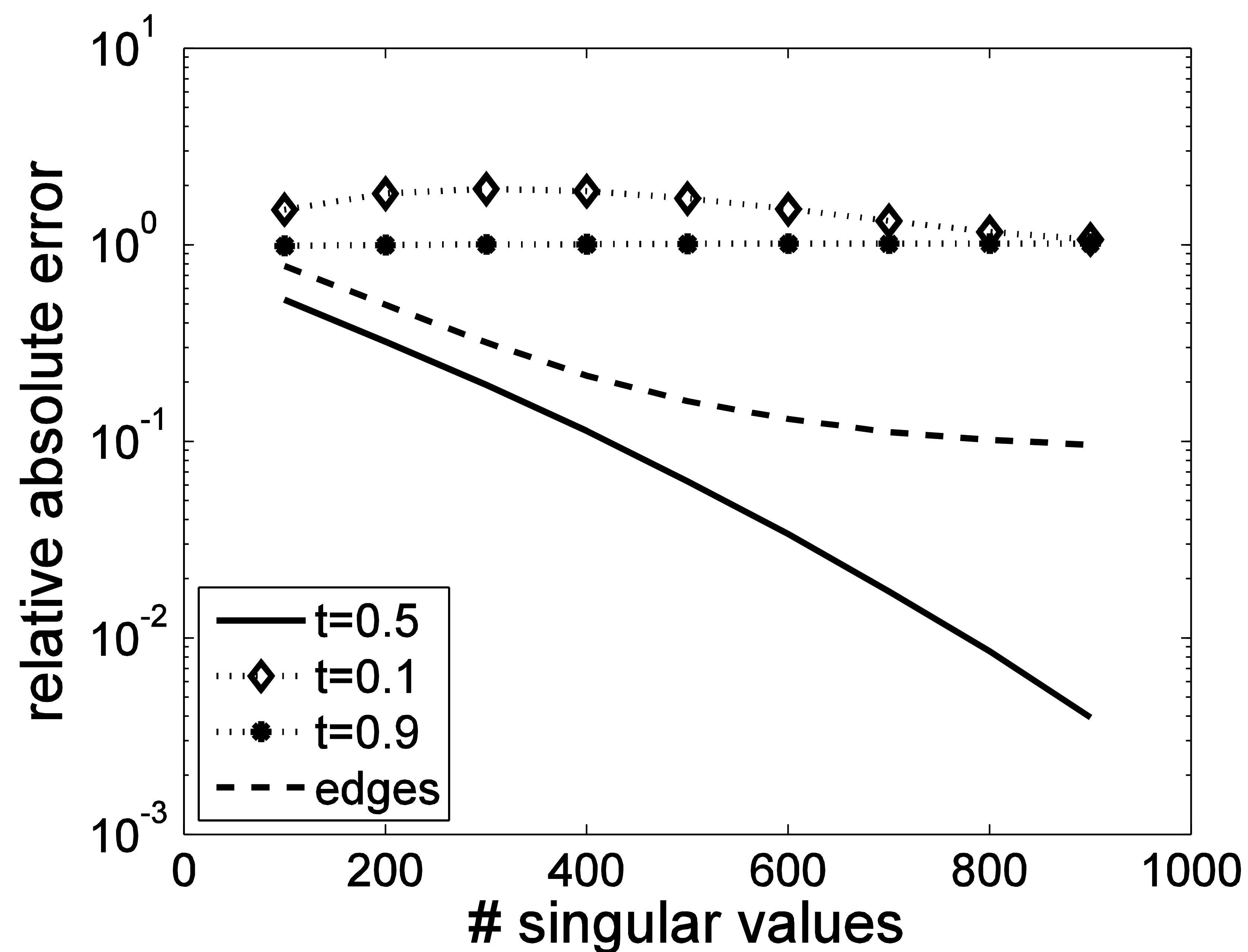
Y axis: relative absolute error (log scale)

relative absolute error:
$$\frac{\|B\hat{M} - M\|}{\|M\|}$$

¹Zheleva *et al.* WWW '09

Experiments

Flickr dataset¹: 2000 users x 1989 groups
Density 5%



X axis: number k of highest magnitude singular values

Y axis: relative absolute error (log scale)

Matrix M has rank 1989, but with only 900 singular values we can achieve almost perfect reconstruction.

relative absolute error:

$$\frac{\|B\hat{M} - M\|}{\|M\|}$$

¹Zheleva *et al.* WWW '09

Outline

Problem definition

Connection to SVD

Greedy SVD reconstruction

Experiments

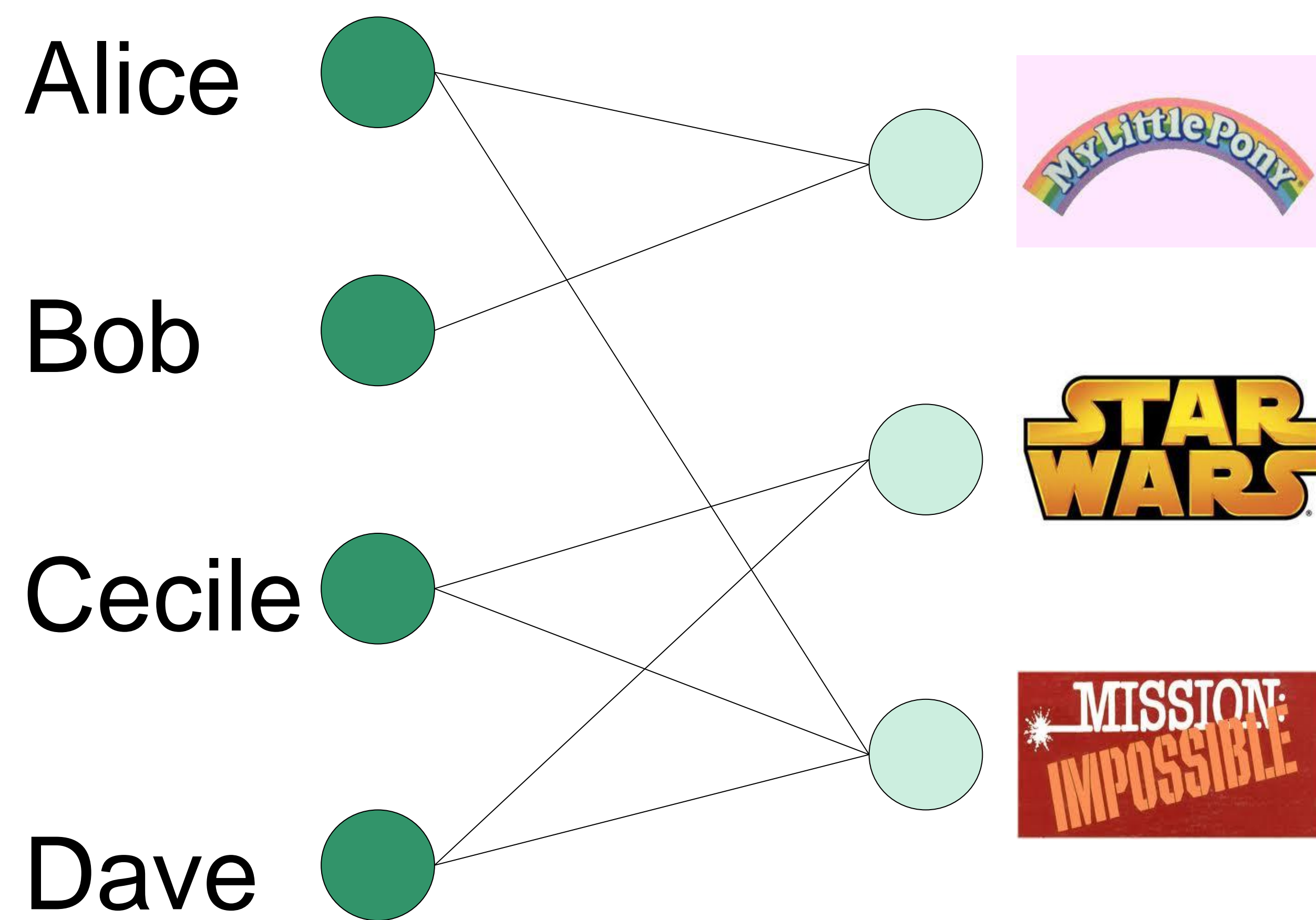
Discussion

	Alice	Bob	Cecile	Dave
Alice	2	1	1	1
Bob	1	1	0	0
Cecile	1	0	2	2
Dave	1	0	2	2

L

			
Alice	1	0	1
Bob	1	0	0
Cecile	0	1	1
Dave	0	1	1

M



			
	2	0	1
	0	2	2
	1	2	3

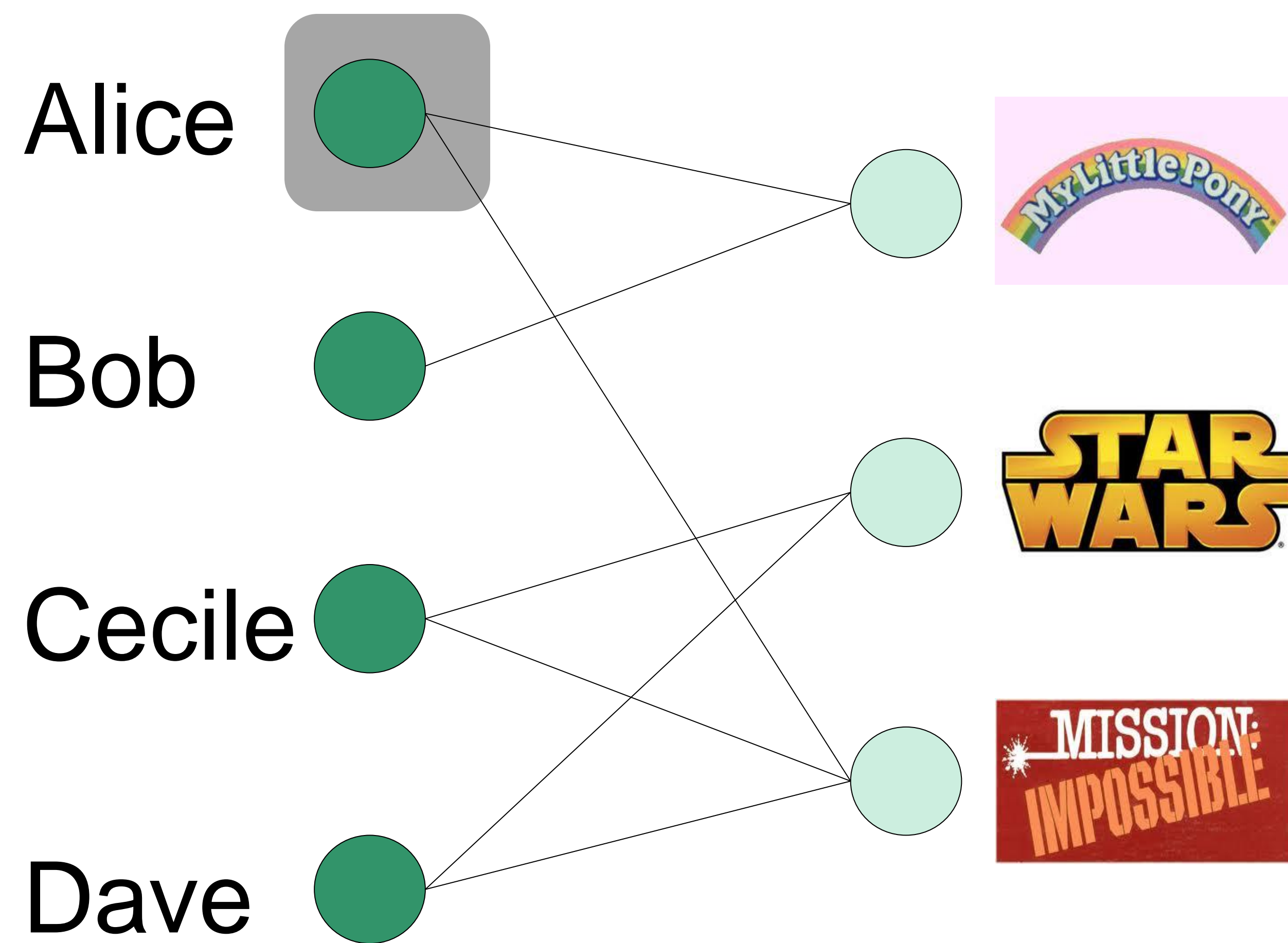
R

	Alice	Bob	Cecile	Dave
Alice	2	1	1	1
Bob	1	1	0	0
Cecile	1	0	2	2
Dave	1	0	2	2

L

			
Alice	1	0	1
Bob	1	0	0
Cecile	0	1	1
Dave	0	1	1

M



			
	2	0	1
	0	2	2
	1	2	3

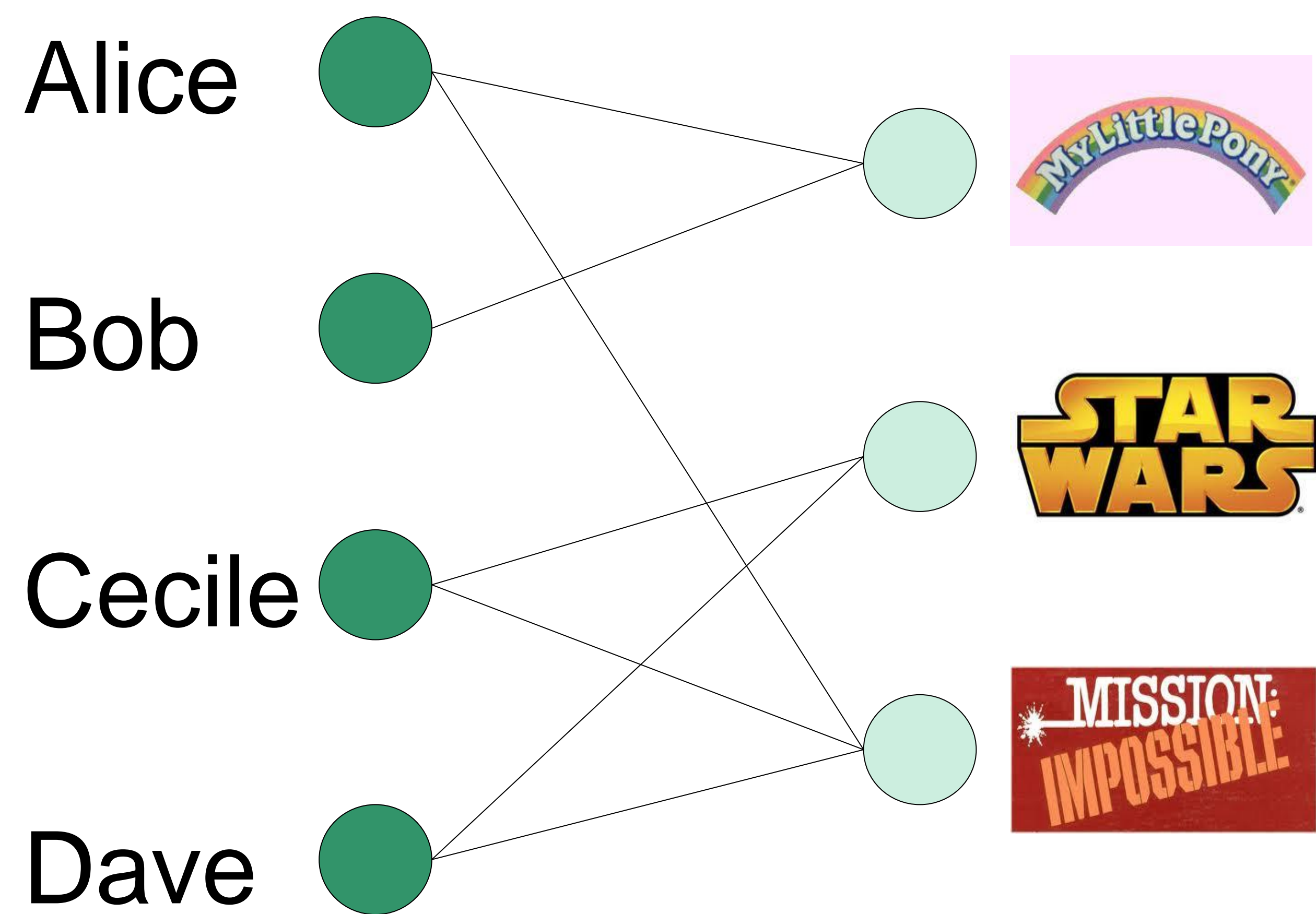
R

	Alice	Bob	Cecile	Dave
Alice	2	1	1	1
Bob	1	1	0	0
Cecile	1	0	2	2
Dave	1	0	2	2

L

			
Alice	1	0	1
Bob	1	0	0
Cecile	0	1	1
Dave	0	1	1

M



			
	2	0	1
	0	2	2
	1	2	3

R

	Alice	Bob	Cecile	Dave
Alice	2	1	1	1
Bob	1	1	0	0
Cecile	1	0	2	2
Dave	1	0	2	2

L

			
Alice	1	0	1
Bob	1	0	0
Cecile	0	1	1
Dave	0	1	1

M

Knowing degrees of nodes carries a lot of extra information

			
	2	0	1
	0	2	2
	1	2	3

R

	Alice	Bob	Cecile	Dave
Alice	0	1	1	1
Bob	1	0	0	0
Cecile	1	0	0	2
Dave	1	0	2	0

L'

			
Alice	1	0	1
Bob	1	0	0
Cecile	0	1	1
Dave	0	1	1

M

Knowing degrees of nodes carries a lot of extra information

Obtain L' and R' by setting the main diagonals to 0.

			
	0	0	1
	0	0	2
	1	2	0

R'

	Alice	Bob	Cecile	Dave
Alice	0	1	1	1
Bob	1	0	0	0
Cecile	1	0	0	2
Dave	1	0	2	0

L'



Knowing degrees of nodes carries a lot of extra information

Reconstruction problem:

Given L' and R' reconstruct M.

	My Little Pony	STAR WARS	MISSION: IMPOSSIBLE
My Little Pony	0	0	1
STAR WARS	0	0	2
MISSION: IMPOSSIBLE	1	2	0

R'

Conclusions

(Bipartite) graphs can be reconstructed from neighborhood data with quite high accuracy.

Often a smaller than $\text{rank}(M)$ number of singular values is sufficient.

Thank You!