Interdependence Between Network Layers

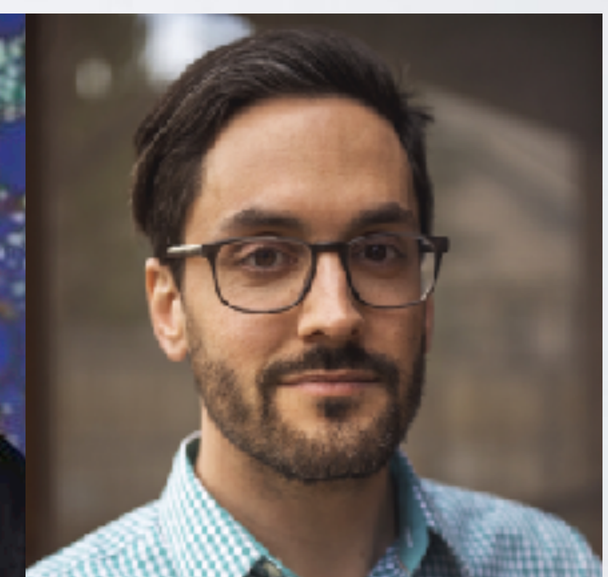
Cristopher Moore, Santa Fe Institute

joint work with

Caterina De Bacco, SFI/Columbia/Max Planck

Daniel Larremore, SFI/Boulder

Eleanor Power, SFI/London School of Economics



Multilayer networks

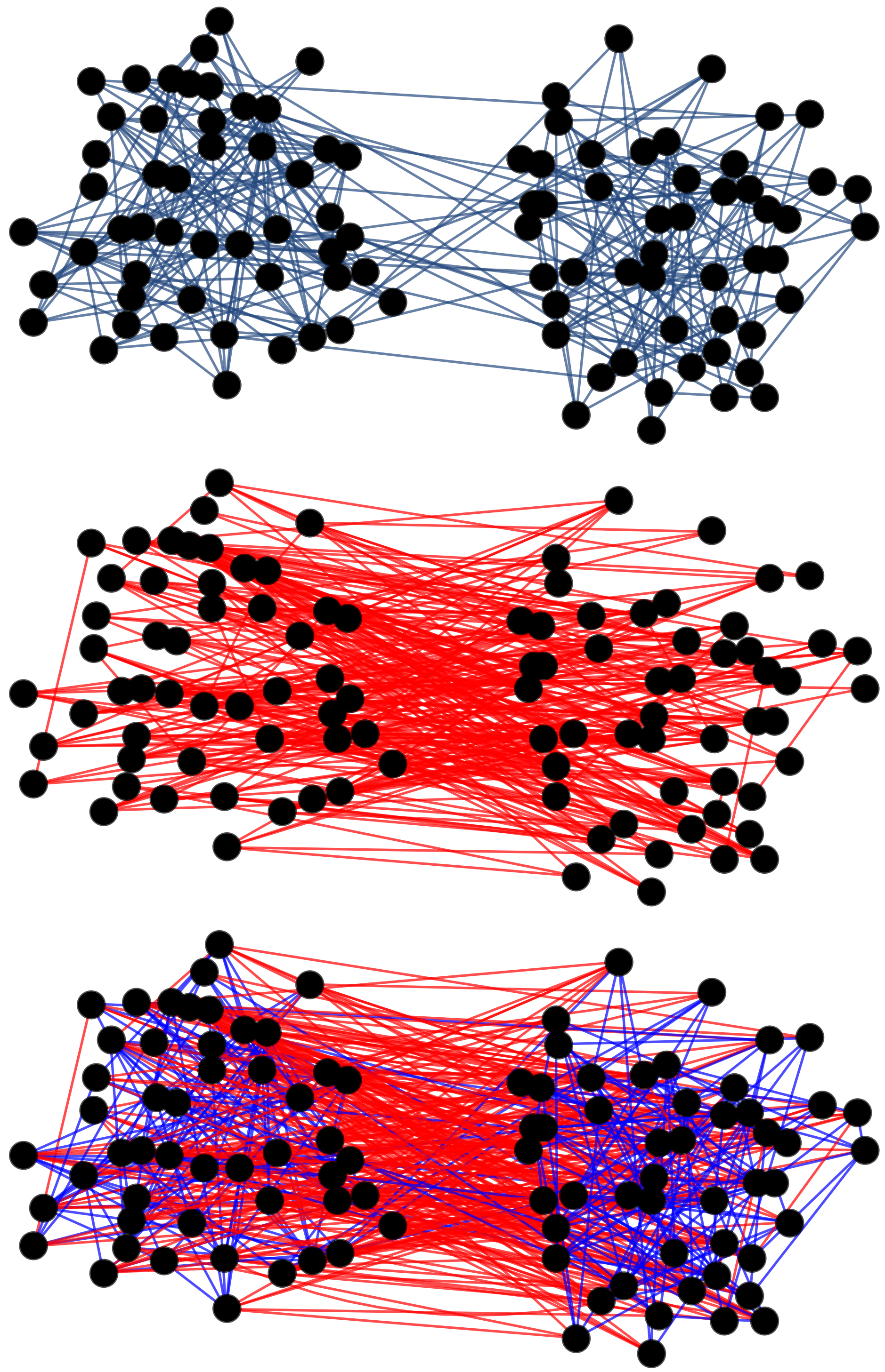
Different types of links: social relationships, modes of transportation, loci of similarity…

Hypothesis: explained by common community structure…

…even if the layers seem very different

Don't ask whether two layers are correlated: ask whether knowing one helps predict the other

Multilayer networks



assortative

disassortative

aggregated

Modeling the adjacency tensor

$$A_{ij}^{(\alpha)} = \#\text{ of links from } i \text{ to } j \text{ in layer } \alpha$$

$$A_{ij}^{(\alpha)} \sim \text{Poisson with mean } M_{ij}^{(\alpha)}$$

# of groups → K

$$M_{ij}^{(\alpha)} = \sum_{k,\ell=1}^{K} u_{ik} v_{j\ell} w_{k\ell}^{(\alpha)}$$

membership of vertex *i* in group *k* → u_{ik}

membership of vertex *j* in group *l* → $v_{j\ell}$

density of edges of type α from group *k* to group *l* → $w_{k\ell}^{(\alpha)}$

Mixed-membership

Allows inhomogeneous degrees

Different roles for incoming and outgoing links

Arbitrary structure in each layer: assortative, core-periphery, etc.

Aside: tensor factorization

rank-one matrix: $A_{ij} = u_i v_j$

rank-*K* matrix: $A_{ij} = \sum_{\ell=1}^{K} u_i^{(\ell)} v_j^{(\ell)}$

to compute matrix rank, just use linear algebra
can approximate (minimize L_2) using singular values

rank-one tensor: $A_{ijk} = u_i v_j w_k$

rank-*K* tensor: $A_{ijk} = \sum_{\ell=1}^{K} u_i^{(\ell)} v_j^{(\ell)} w_k^{(\ell)}$

tensor rank is NP-complete!
rank of matrix multiplication tensor is unknown
even minimizing L_2 requires an iterative algorithm

The EM algorithm

$\rho_{ijkl}^{(\alpha)}$ is our estimate of the probability that an edge $i \to j$ in layer α is due to their being in groups *k* and *l*

maximizing the likelihood gives update equations:

$$u_{ik} = \frac{\sum_{j,\alpha} A_{ij}^{(\alpha)} \sum_{\ell} \rho_{ijk\ell}^{(\alpha)}}{\sum_{\ell} \left(\sum_j v_{j\ell}\right) \left(\sum_{\alpha} w_{k\ell}^{(\alpha)}\right)} \qquad v_{j\ell} = \frac{\sum_{i,\alpha} A_{ij}^{(\alpha)} \sum_{k} \rho_{ijk\ell}^{(\alpha)}}{\sum_{k} \left(\sum_i u_{ik}\right) \left(\sum_{\alpha} w_{k\ell}^{(\alpha)}\right)}$$

$$w_{k\ell}^{(\alpha)} = \frac{\sum_{ij} A_{ij}^{(\alpha)} \rho_{ijk\ell}^{(\alpha)}}{\left(\sum_i u_{ik}\right) \left(\sum_j v_{j\ell}\right)} \qquad \rho_{ijk\ell}^{(\alpha)} = \frac{u_{ik} v_{j\ell} w_{k\ell}^{(\alpha)}}{\sum_{k'\ell'} u_{ik'} v_{j\ell'} w_{k'\ell'}^{(\alpha)}}$$

iterate until fixed point = local maximum of the likelihood

(gives a point estimate: there are also Bayesian versions)

Cross-validation and link prediction

Can the model fill in missing data?

Hide 20% of the links, use 80% as training data

Generative models assign probabilities to missing links: use AUC to measure accuracy

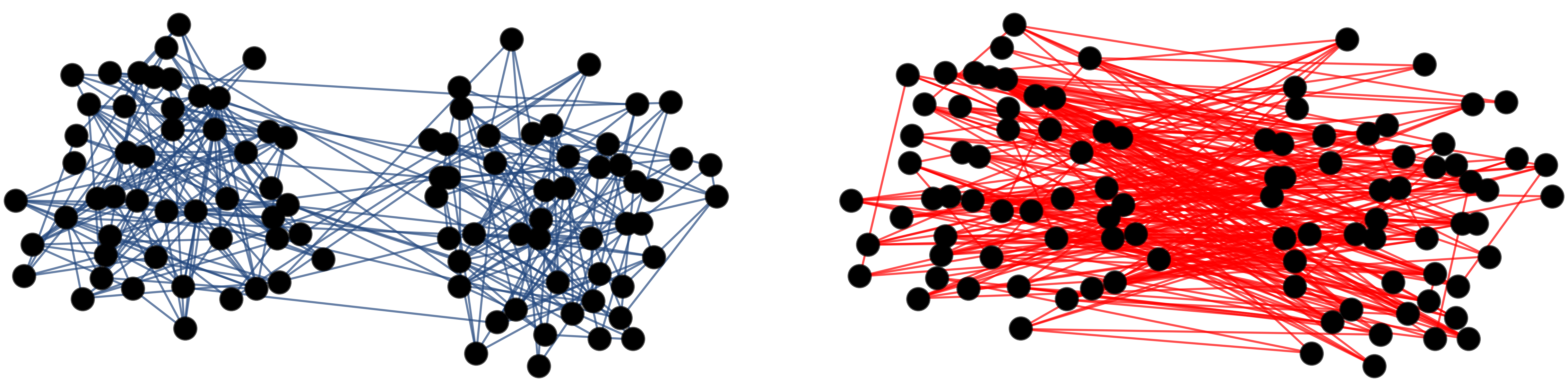
Avoid overfitting, and select # of communities

But! "Best" model depends on what kind of data is hidden, and what we are trying to predict: application-dependent (as it should be)

Using link prediction to measure interdependence between layers

Does knowing one layer help predict links in another one?

"Similar" (not correlated!) layers have common community structure, and help predict each other



Using link prediction to measure interdependence between layers

Does knowing one layer help predict links in another one?

“Similar” (not correlated!) layers have common community structure, and help predict each other

Layers are redundant if they reveal same latent features of the nodes: don’t need to ask about both

But knowing one layer may make it harder to predict another, if their structures are inconsistent

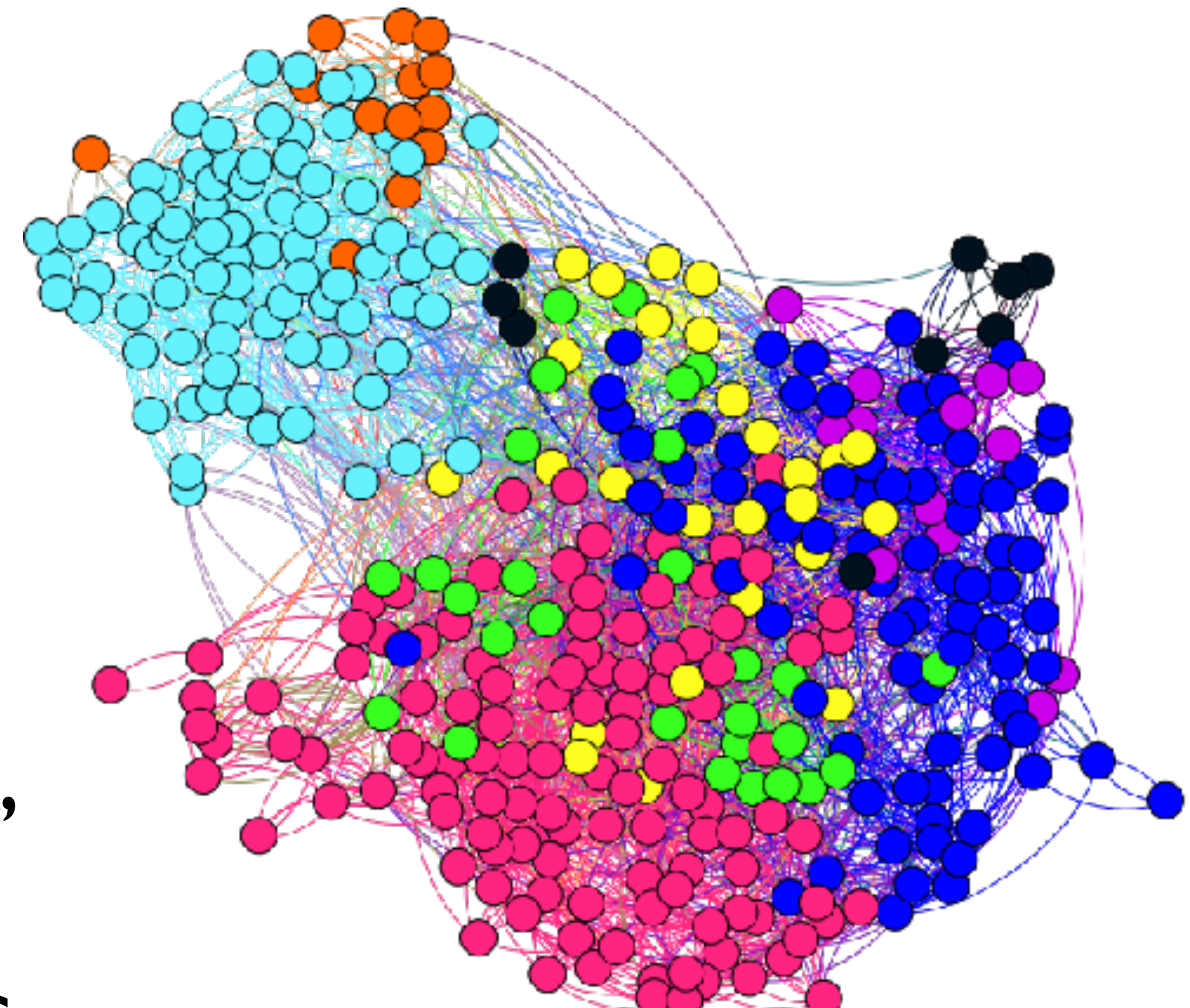
Experiments on social networks

Two Indian villages

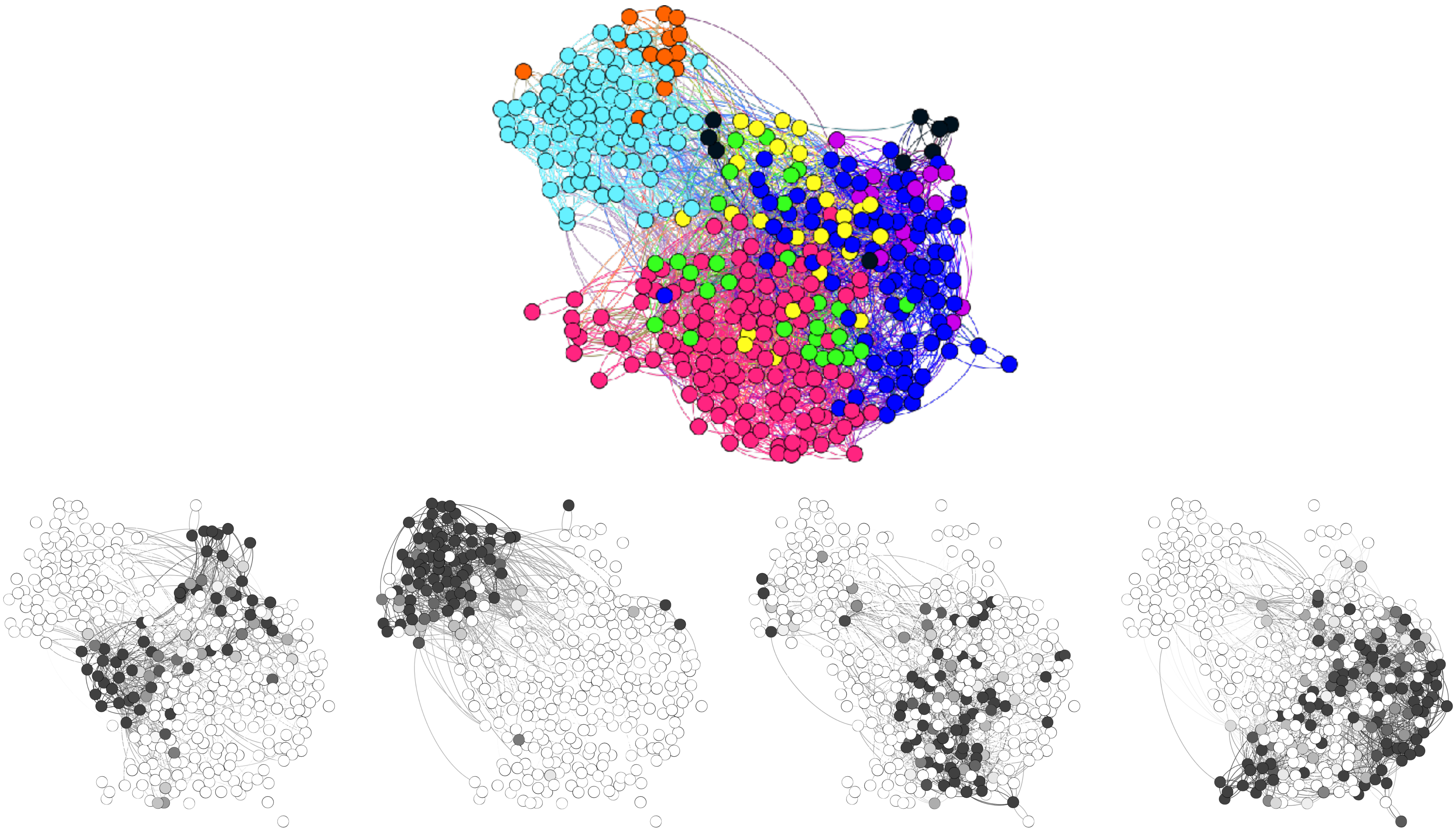
~400 nodes, ~7000 edges

One village consists of two separate hamlets

12 layers: looking for work, babysitting, borrowing, discussing important issues

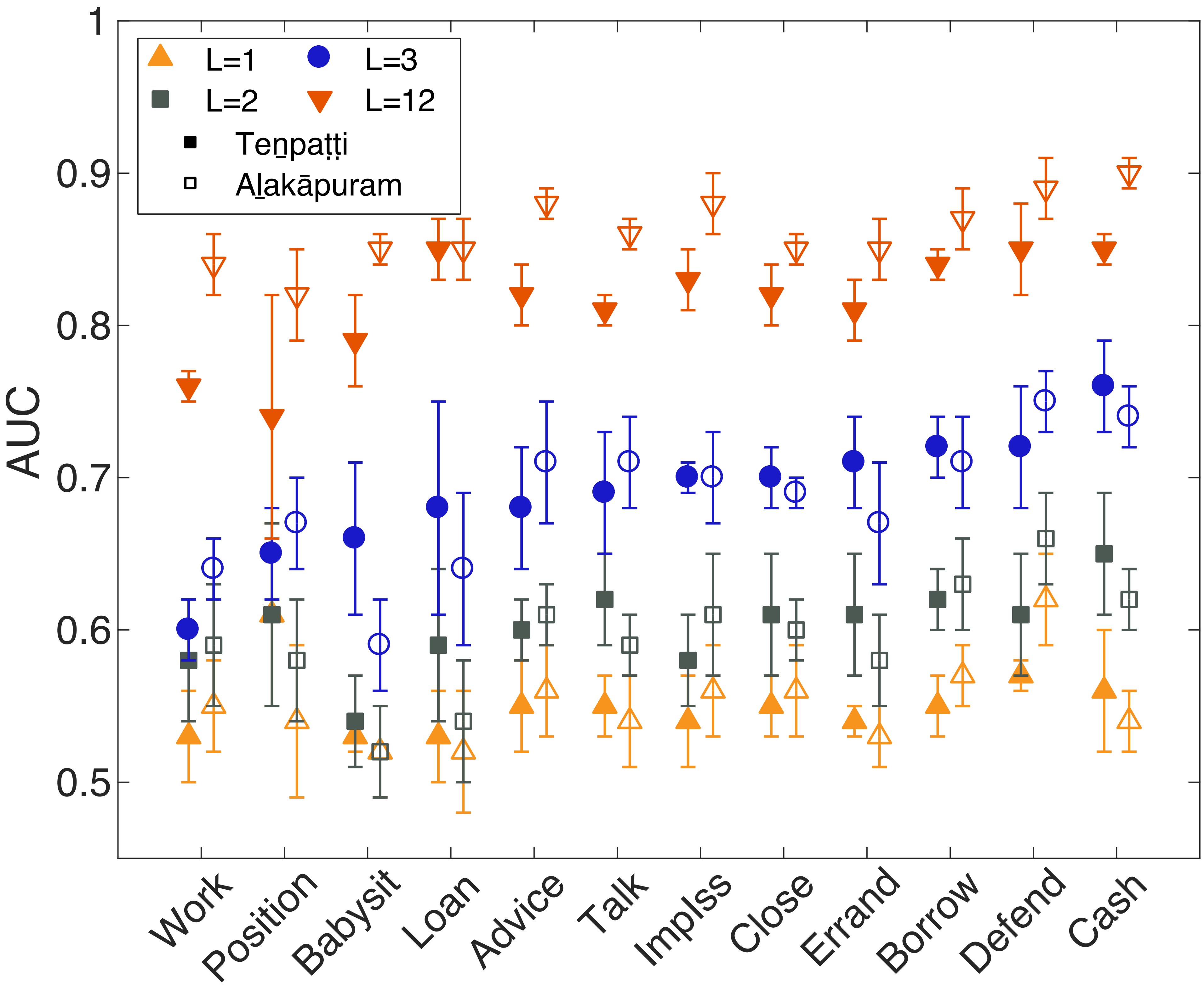


A common community structure



Correlated with caste, gender, and geography

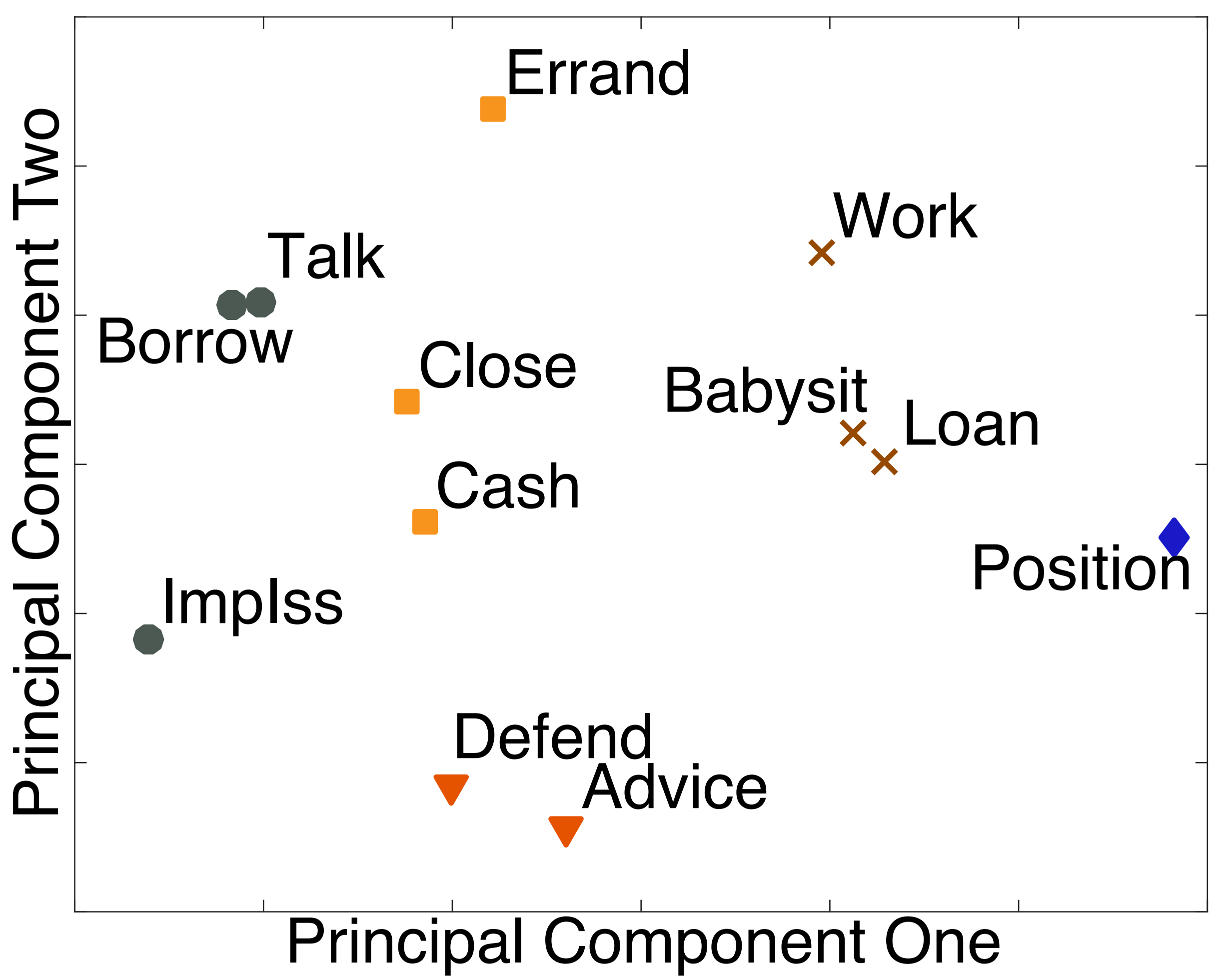
Knowing more layers helps

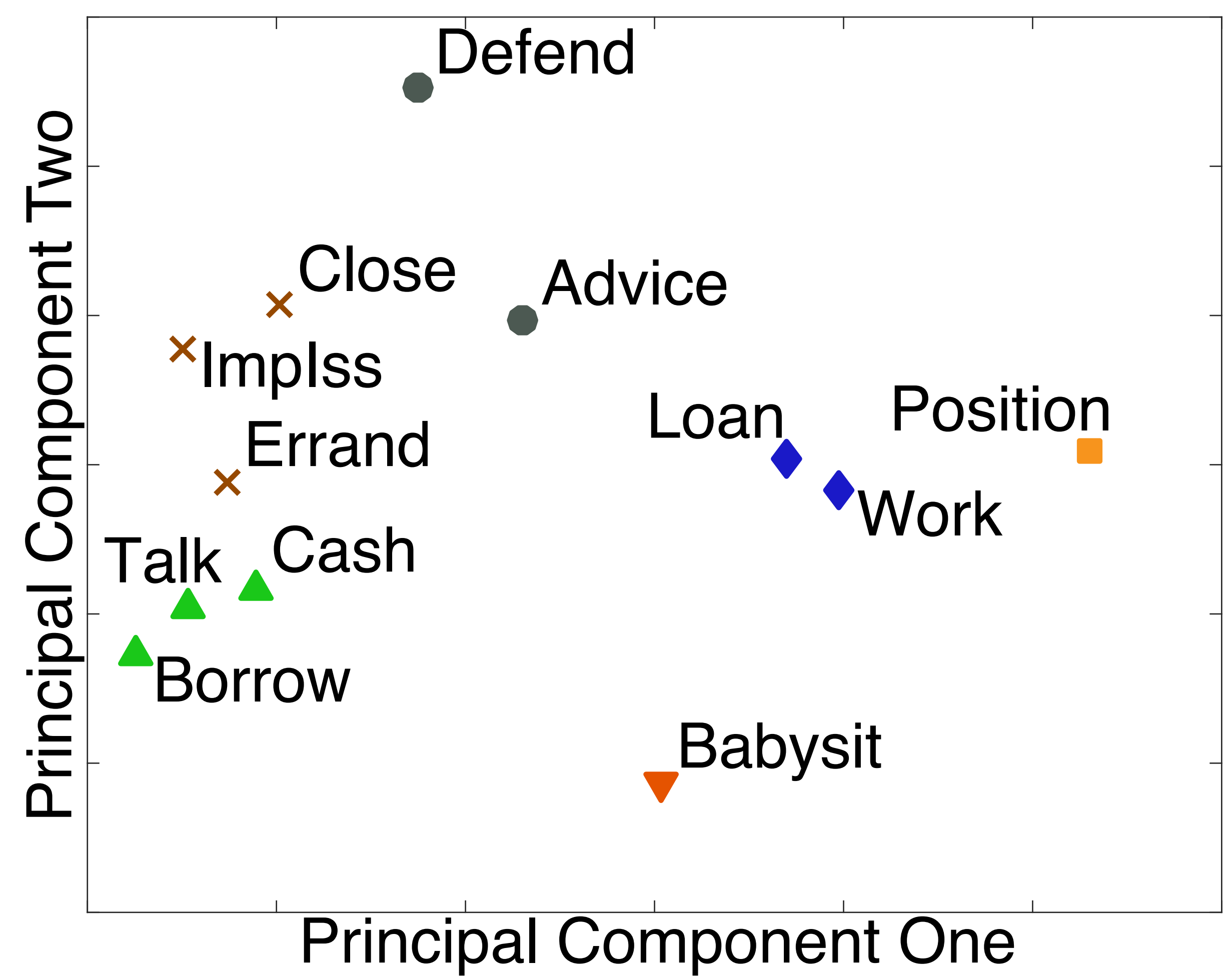


Interdependence and similarity

"Who do you discuss important issues with?" helps predict many layers; looking for work, babysitting less so

Can also cluster layers by affinity matrices $w^{(\alpha)}$:



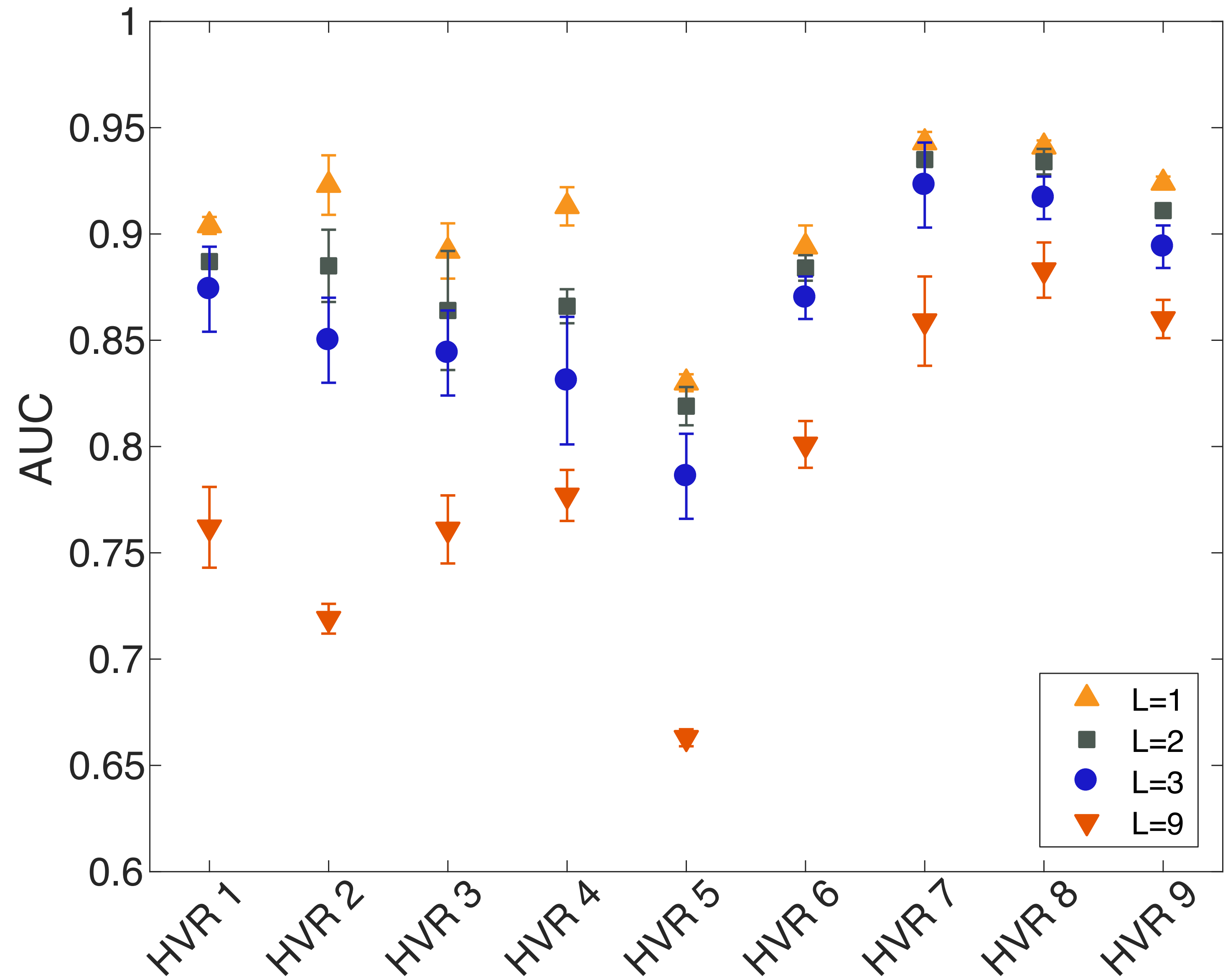


In contrast: malaria genetics

~300 variants of a gene in the malaria parasite

9 layers, similarity at different loci

highly variable to avoid immune response



more layers, less accurate: no consistent communities

the joy of negative results…

Conclusions and questions

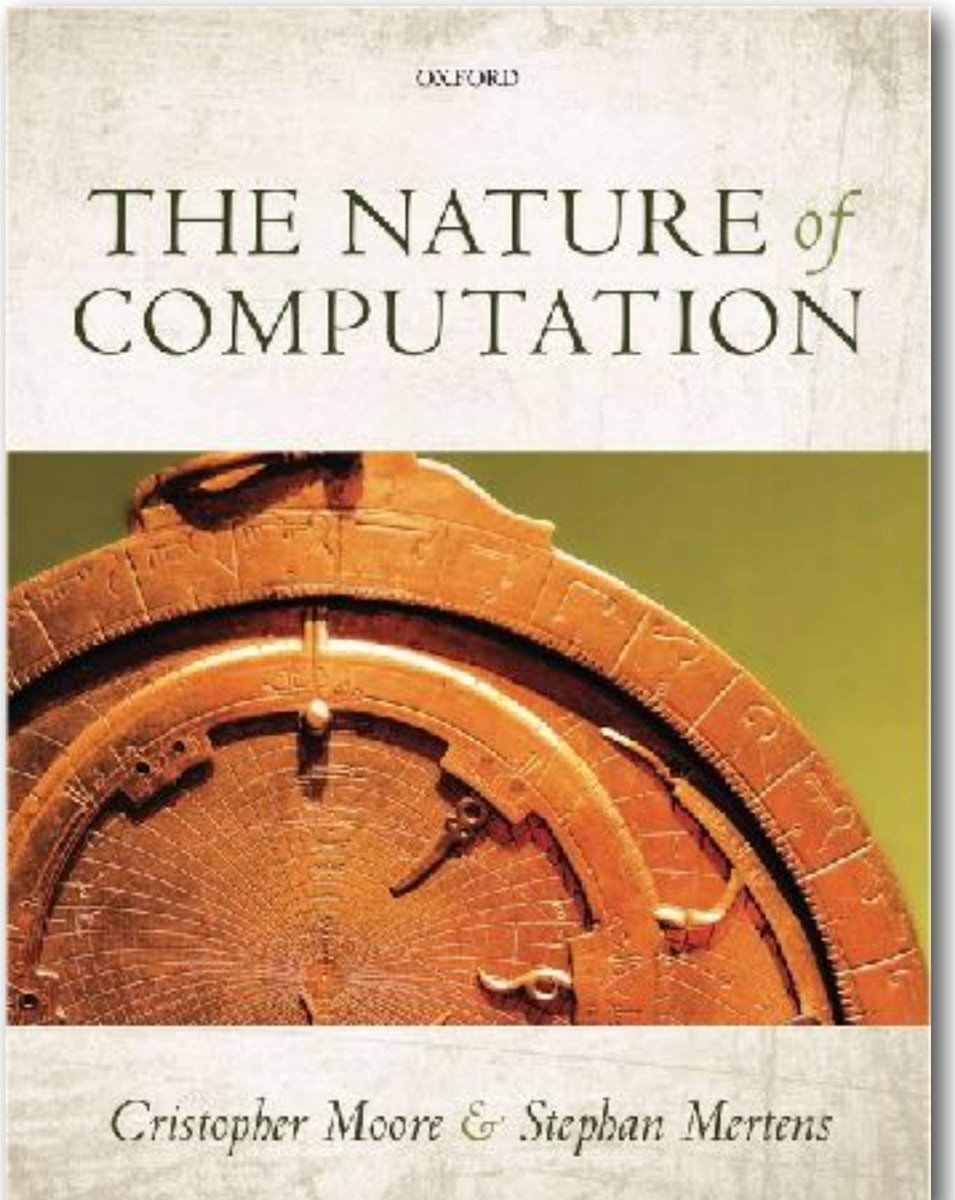
Can use tensor factorization to express models of multilayer networks

Instead of correlation, ask whether one layer helps predict another: do they reveal similar latent structures?

This is model-dependent… what are we missing?

Local rules, e.g. if (a,b) and (b,c) have relation #1, then (a,c) have relation #2

Shameless plug



www.nature-of-computation.org

To put it bluntly: this book rocks! It somehow manages to combine the fun of a popular book with the intellectual heft of a textbook.

Scott Aaronson, MIT

This is, simply put, the best-written book on the theory of computation I have ever read; one of the best-written mathematical books I have ever read, period.

Cosma Shalizi, Carnegie Mellon