Testing Network Hypotheses

Statistical Issues

- Samples non-random
- Often work with populations
- Observations not independent
- Distributions unknown

Logic of Permutation Test

- Compute test statistic
  - e.g., correlation or difference in means
  - Correlation between centrality and salary is 0.384 or difference in mean centrality between the boys and the girls is 4.95.
  - Ask what are the chances of getting such a large correlation or such a large difference in means if the variables are actually completely independent?

Logic of Permutation Test

- So to evaluate an observed correlation between two variables of 0.384, we want to
  - correlate thousands of variables similar to the ones we are testing that we know are truly independent of each other, and
  - see how often these independent variables are correlated at a level as large as 0.384
    - The proportion of random correlations as large observed value is the p-value of the test
- How to obtain thousands of independent variables whose values are assigned independently of each other?
  - Fill them with random values
    - But need to match distribution of values
  - Permute values of one with respect to the other

Outline of Permutation Test

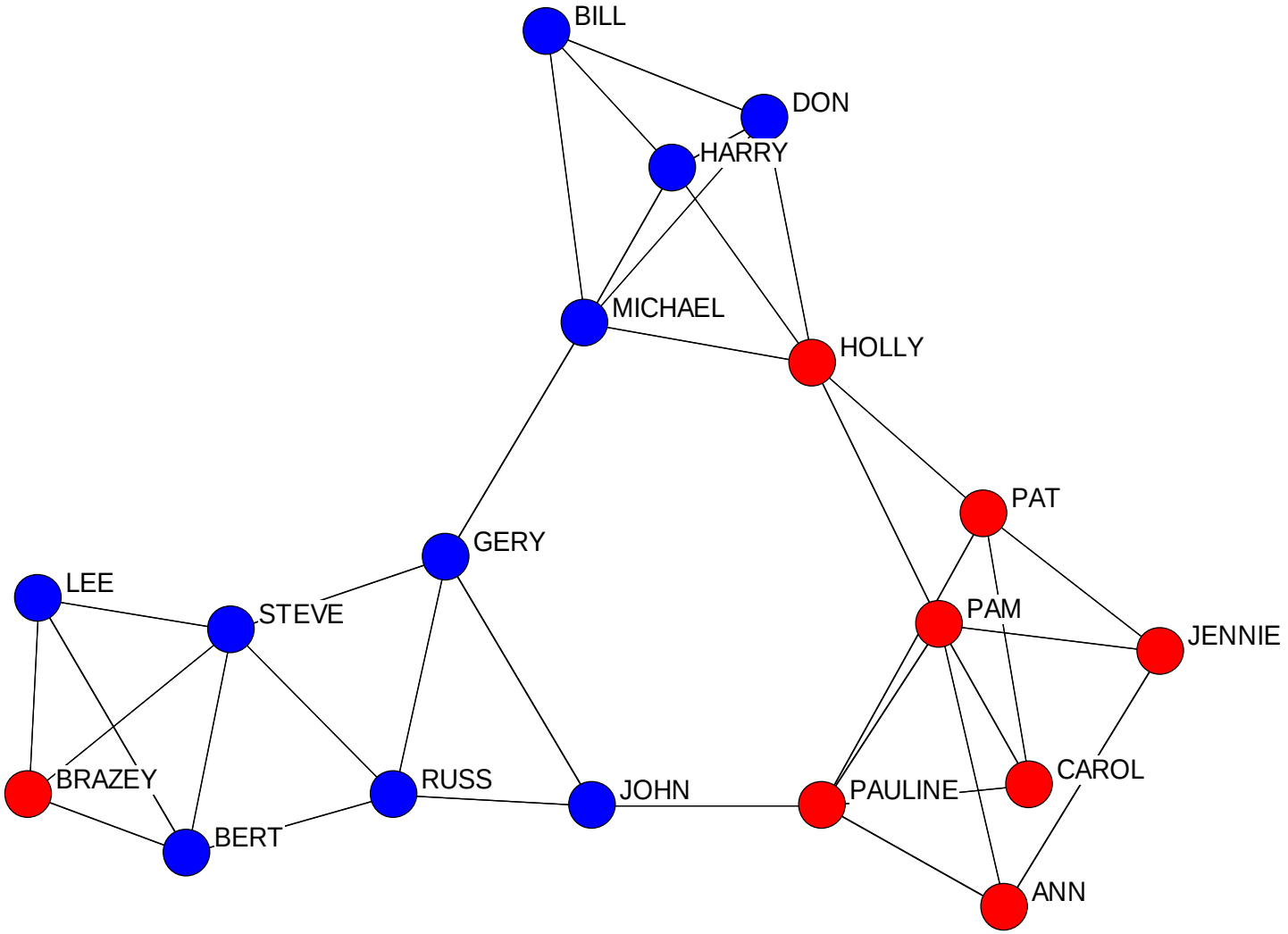
- Get observed test statistic
- Construct a distribution of test statistics under null hypothesis
  - Thousands of permutations of actual data
- Count proportion of statistics on permuted data that are as large as the observed
  - This is the p-value of the test

Note

- Can use collected network data
- Can use hypothesized structural data
- Can use correlation, regression plus many other statistical routines. Good examples are in Hanneman book.

E-I index

- Network data plus a partition
- E= external links
- I= internallinks
- Index = (E-I)/(E+I)
- Permutation test used in UCINET
- Reports made at network, group and individual level.


BILL
DON
HARRY
MICHAEL
HOLLY
PAT
GERY
LEE
STEVE
PAM
JENNIE
BRAZEY
RUSS
JOHN
PAULINE
CAROL
BERT
ANN