Dr. Ali Tafti
IDS 564: Social Media and Network Analysis, Fall 2018
Lab 5: Community detection, and class readings EK, ch. 19.
Recommended reading: KC, ch. 7

EK: Easley, David, and Jon Kleinberg. *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press, 2010.

KC: Kolaczyk, Eric D., and Gábor Csárdi. *Statistical analysis of network data with R*. Vol. 65. Springer, 2014. Available online at the UIC library: http://HZ9PJ6FE4T.search.serialssolutions.com/?V=1.0&L=HZ9PJ6FE4T&S=JCs&C=TC0001247657&T=marc&tab=BOOKS

Preparation: Please review the main content in EK chapter 19. Please also study and work through the posted R script, based on chapter 7 of the KC book; the relevant sections of the script are posted with this assignment. You may also find it helpful to read the chapter 7 of the KC book. Use R help functions to understand the functions in that script. The lab requires you to understand the functionality just well enough to make some incremental additions to answer some questions.

The script first uses the French blogs dataset to show a viola plot illustrating the number of common neighbors among pairs of nodes that share or do not share an edge. Next, the script uses a dataset based on a series of experiments that capture correlations in activity among the genes in the E Coli bacteria. It uses those correlations to predict links in the network, and compares those predictions to actual links that have been confirmed in prior research. Next, please answer the quiz questions. One of the questions refers to Fig. 19.27 from the EC book, which is reproduced here. Please submit your quiz answers prior to the hard deadline. Please submit your final R script and a Word or PDF file with your viola plots, network plots, and any additional analysis used to answer the quiz questions.

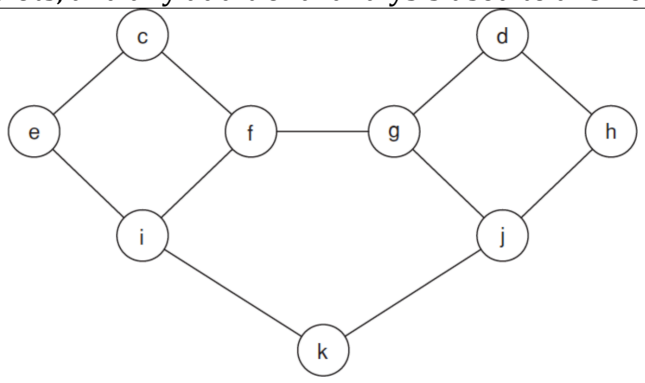

Figure 19.27.