Chapter 1

Supplement to Chapter 1: Guidelines for Projects

John R. Jungck and Rama Viswanathan

See data files and manual for BioGrapher, which are accessible from the book's companion website.

Project 1.1. Two beautiful examples that would lend themselves to extending the interval graphs to handle circular arc graphs (36) are shown in Figure 1.11. □

Project 1.2. Another appropriate example is based on data in Breyne et al. [39]. □

While the analysis of each network in both these projects could become a full-fledged research project, we suggest that you begin by analyzing these networks with the concepts described above.

- Develop an adjacency matrix for each network.
- Construct intersection and complementary graphs and visualize them in the three layouts—spring, radial, and circular—available in the Excel BioGrapher workbook.
- Identify maximal cliques.
- Check for Z_4's.

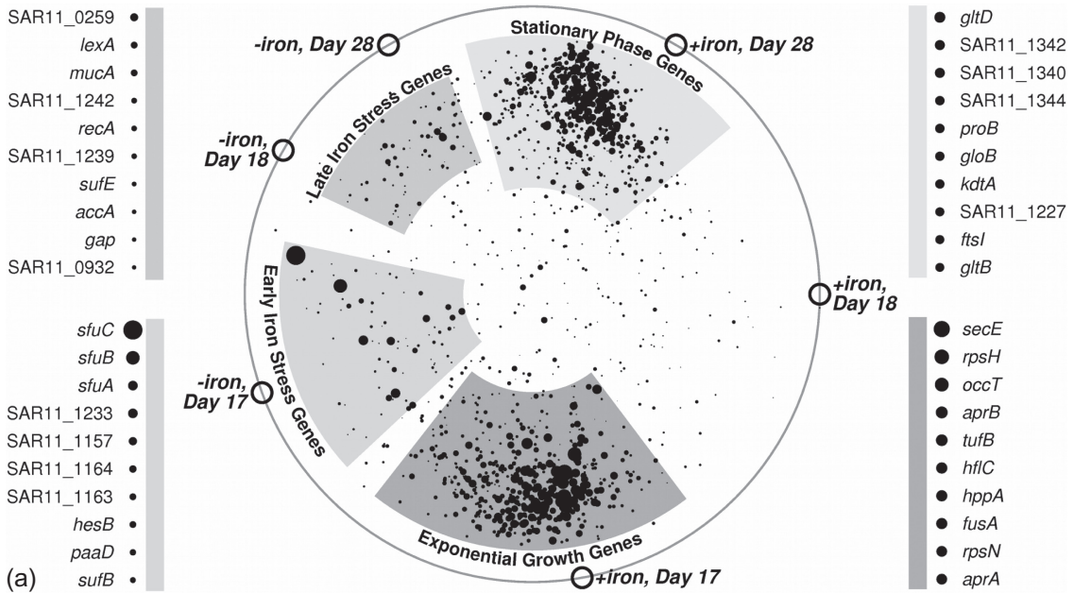
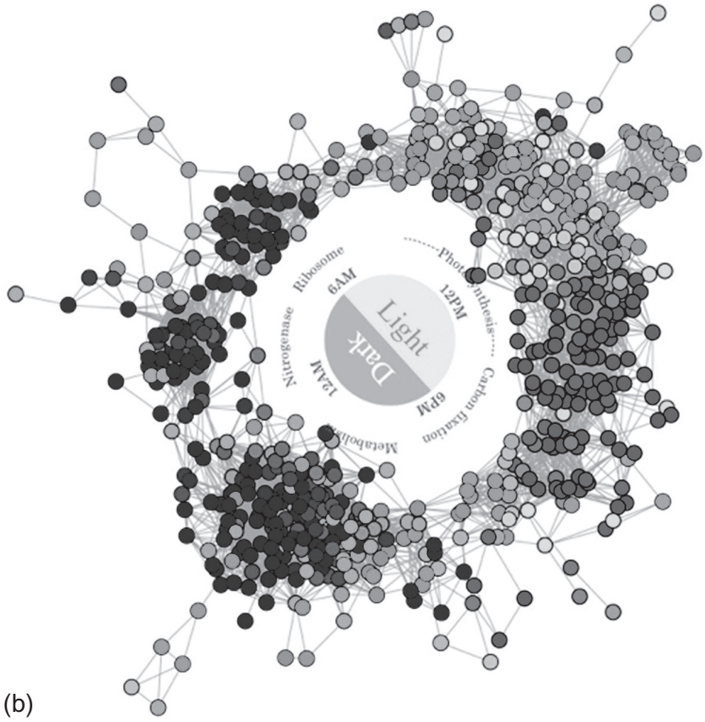


FIGURE 1.14 Examples of circular arc graphs. (a) Transcriptional and translational regulatory responses to iron limitation in the globally distributed marine bacterium *Candidatus Pelagibacter ubique* [38].

Algebraic and Discrete Mathematical Methods for Modern Biology. http://dx.doi.org/10.1016/B978-0-12-801213-0.09987-1
Copyright © 2015 Elsevier Inc. All rights reserved.



(b)

FIGURE 1.14 (b) A model of cyclic transcriptomic behavior in *Cyanothece* species ATCC 51142 [39].

- Determine whether the complementary graphs satisfy the transitive orientability property to construct a comparability graph.
- Export the adjacency matrix for each network into javaBENZER and re-arrange into Shkurba form as much as possible.
- Check whether you can construct something close to an interval graph or a circular arc graph from the above elements.
- Look for a causal biological relationship for the maximal cliques, their connections and orders such as time or location of expression, collisions or binding between molecules, metabolic flux, and signal transmission. Each of the three networks has a substantial different biological context. In Figure 1.14a, Smith et al. [37] examined responses to iron limitation in a globally distributed marine bacterium by measuring transcriptional (reading of DNA into messenger RNA) and translational (reading of messenger RNA to construct protein sequences). In Figure 1.14b [38], "two normally incompatible activities fixing nitrogen in the dark and photosynthesizing in the light" were examined in an "ocean-dwelling cyanobacteria that is capable of it mak[ing] this switch inside the same cell every light/dark cycle (normally about 12 hours). This makes it interesting from the standpoint of bioenergy …production [as well as its] regulation [because] the process of how it is able to drastically rearrange it's machinery every 12 hours is not well understood." The authors [38] made "measurements of levels of gene expression" and were surprised that the circular clock of co-expressions of clusters of genes corresponded so well to phases in the light/dark cycle. Breyne et al. [39] did similar analyses of gene expression during cell division in plants. It is important to check the biological context of how each of these networks in order to relate the topological orderings to underlying causal relationships among physical materials.
- Consider whether the network is resilient to perturbation (particularly useful if investigators have conducted "knock out" experiments which are equivalent to removing vertices from a network or if you are examining the evolutionary stable strategy for robustness in the face of perturbation).

1.1 GUIDING QUESTIONS FOR PROJECT 1.1

In your analysis of the data in Figure 1.14a, Smith et al. [37], you should look for congruences between maximal cliques illustrated in javaBENZER or BioGrapher with the four colored (dark blue, yellow, green, and light blue) regions in the original figure. First: Are the partitions between colored clusters as the authors chose? Many vertices are loosely connected to these clusters so the authors have done some sort of topological reduction of the complexity of the data. Are you comfortable with their choices? If so, why? If not, why not? Have you produced clusters that make better sense to you? If so, are you a lumper or a splitter? Second: The data are arranged in a circle in Figure 1.14a. Do you find that the data closely fit to a circular arc graph? Was the circular layout in BioGrapher better than the spring or radial configurations in displaying associations of the data? If you permutated your javaBENZER matrix, were you finding linkage between the upper left and the lower right cliques? If not, since the 3 O'clock portion of the clock claimed by the authors in Figure 1.14a is so sparse, might not an interval graph representation account for the data just as well without forcing them onto a circular clock? Third: The authors have focused on examining the messenger RNA molecules and protein sequences within these organisms, but how might you make inferences about the microbial communities and their impact on the environment?

In Figure 1.14b [38], the authors made "measurements of levels of gene expression" and were surprised that the circular clock of co-expressions of clusters of genes corresponded so well to phases in the light/dark cycle. AGAIN: in your analysis of the data in Figure 1.14b, McDermott et al. [38], you should look for congruences between maximal cliques illustrated in javaBENZER or BioGrapher with the five colored (blue, red, green, yellow, and purple) regions in the original figure. First: Are the partitions between colored clusters as the authors chose? Many vertices are loosely connected to these clusters so the authors have done some sort of topological reduction of the complexity of the data. Are you comfortable with their choices? If so, why? If not, why not? Have you produced clusters that make better sense to you? If so, are you a lumper or a splitter? Notice that the clusters in Figure 1.14b are more tightly packed than in Figure 1.14a, but that they also overlap substantially—are you able to more cleanly separate the clusters? Second: The data are arranged in a circle in Figure 1.14b. Do you find that the data closely fit to a circular arc graph? Was the circular layout in BioGrapher

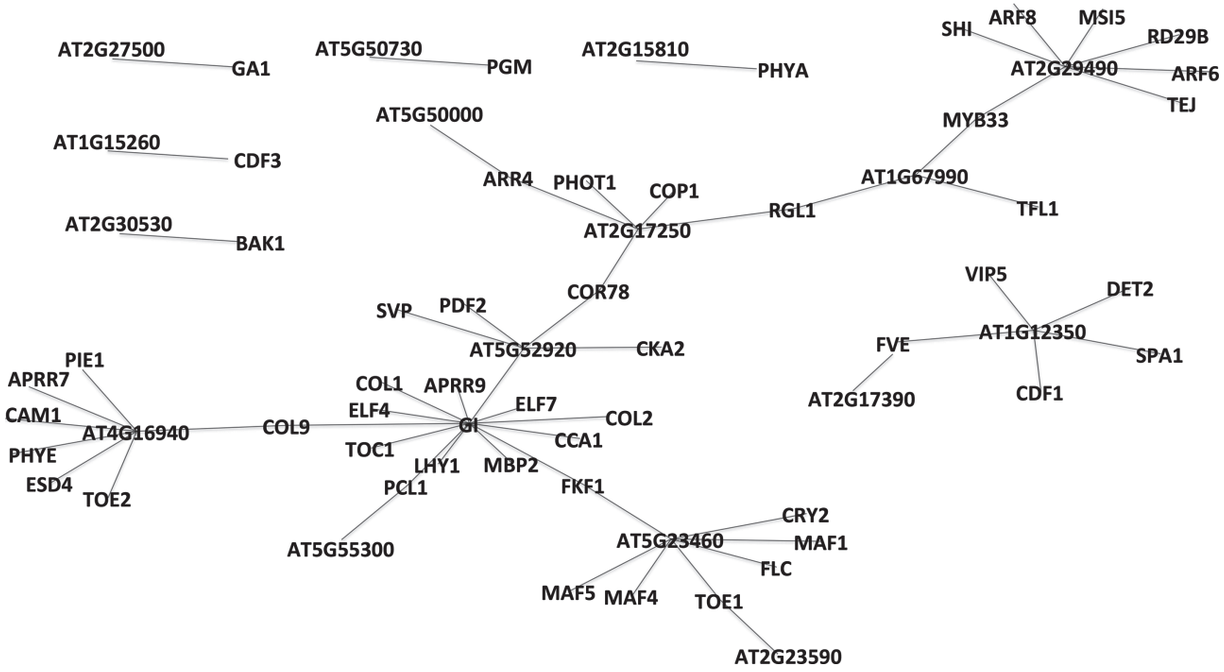


FIGURE 1.18 BioGrapher spring layout visualization of the *Arabidopsis* gene data set published recently by Keurentjes et al. [42]. This graph is a tree and has some high-degree hubs.

better than the spring or radial configurations in displaying associations of the data? If you permutated your javaBENZER matrix, were you finding linkage between the upper left and the lower right cliques? Third, look at the supplementary data for the McDermott et al. [38] article: http://jasonya.com/wp/wp-content/uploads/2013/03/SupplementalInformation_04_18_2011.pdf. They present statistics on six different clustering approaches. How do these various approaches compare to your constructing of clusters? Fourth, they discuss the important biological phenomenon: "two normally incompatible activities fixing nitrogen in the dark and photosynthesizing in the light" were examined in an "ocean-dwelling cyanobacteria that is capable of it mak[ing] this switch inside the same cell every light/dark cycle (normally about 12 hours). This makes it interesting from the standpoint of bioenergy …production [as well as its] regulation [because] the process of how it is able to drastically rearrange it's machinery every 12 hours is not well understood." The authors have divided the clock into five different areas with five different colors, what if you just looked at day versus night or some other binary division? If we ignore the green vertices that are spread all across the network, does a day-night division more cleanly separate the purple and blue clusters from red and yellow clusters?

1.2 GUIDING QUESTIONS FOR PROJECT 1.2

Breyne et al. [39] did similar analyses of gene expression during cell division in plants. It is important to check the biological context of how each of these networks in order to relate the topological orderings to underlying causal relationships among physical materials. We leave it to the reader to translate the above suggestions to this third example.

Recall that in all three cases above, we asked you to: "Consider whether the network is resilient to perturbation (particularly useful if investigators have conducted 'knock out' experiments which are equivalent to removing vertices from a network or if you are examining the evolutionary stable strategy for robustness in the face of perturbation)."

Project 1.3. There are many hub-like networks published in the primary literature. These do not all share the obfuscation critiqued above. For example, the network in Figure 1.18 clearly demonstrates a chain of hubs. □

Keurentjes et al. [42] came to an important conclusion based upon their analysis of this network: "Distant gene expression regulation occurs more frequently but local regulation is stronger." Their inference is that *cis*-regulation of closely linked genes on a chromosome is more important than *trans*-regulation of distantly separated genes or genes on different chromosomes. We invite you to explore further their data set of 176 of the 192 genes known to be involved in flowering in *Arabidopsis*. In particular, we want to draw your attention to a graph theoretic property of the network visualization based on their data (Figure 1.18), namely, that it is a tree and hence a planar graph. Some vertices have a high degree (the hubs) but most vertices only have one neighbor (called spokes). Note that no maximal clique in this graph is larger than 2. Now re-examine Figure 1.17c: while there are a few overlapping edges that cannot be re-arranged into a planar form, the overall appearance of the graph shares much of the hub and spoke aspects of Figure 1.18. Therefore, we might ask: How tree-like is a network without many maximal cliques within it? That is, if we were to eliminate a few edges or collapse small polygons (triangles, squares, or maximal cliques) to form vertices, would the transformed network now be a tree? The topological reduction of graphs is frequently useful for finding linear or circular associations that we previously highlighted with interval and circular arc graphs.

1.3 GUIDING QUESTIONS FOR PROJECT 1.3

The data matrix for the tree shown in Figure 1.17c is given in the supplementary files. Construct and re-arrange the javaBENZER adjacency matrix from this adjacency matrix. First, how do hubs look different from maximal cliques? Second: How does the re-arranged (Shkurba-like) form of a tree look different from that of an interval graph? Third, compare Keurentjes et al.'s [42] approach with that of Neapolitan et al. [40]. In particular, look at their networks in the three layouts in BioGrapher: radial, circular, and spring. Which format best allows you to see clusters (either cliques or hubs) in each group's data? Finally: the authors discuss local versus global properties of their networks. How do the diameters of these graphs compare with those generated in the networks of Projects 1.1 and 1.2 above? How does this compare to the kind of regulation in each system?