Octave routines for network analysis

GB

September 23, 2012

Contents

0 Basic network routines

0.1 Basic network theory

A **graph** is a set of nodes, and an associated set of links between them.

Networks are instantiations of graphs. They often represent real world systems that can be modeled as a set of connected entities.

Network theory is a modern branch of **graph theory**, concerned with statistics on practical instances of mathematical graphs. Graph theory and network theory references are abundant. Social science is probably the most recent instigator of the trend to see the world as a network. In 1967, Milgram conducted his famous small world experiment [1], and found that Omahans are on average six steps away by acquaintance from Bostonians. Other prominent first sources are Price's work on the graph of scientific citations in 1965 [2] and in 1998, Watts and Strogatz's paper on dynamics of small-world networks [3].

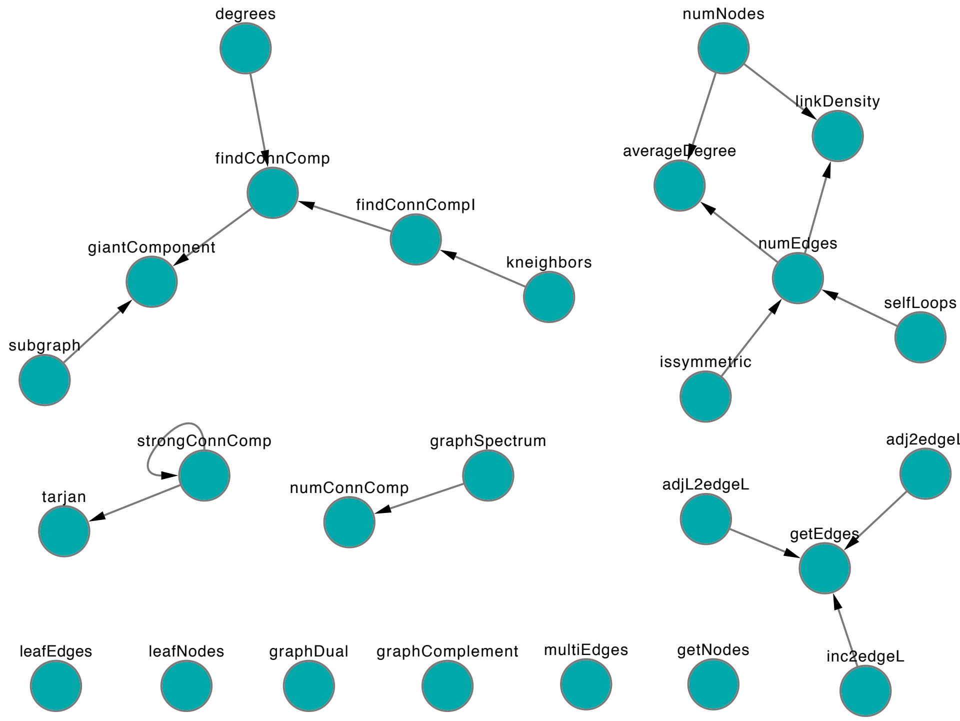

Figure 1: Graph of functions in this toolbox. An edge points from routine A to routine B if routine A is used within routine B.

Nowadays, there is no shortage of books and reviews on networks. Below is a non-exhaustive list of good reads [4] [5] [6] [7].

- S. Wasserman and K. Faust, Social network analysis, Cambridge University Press, 1994
- Duncan J. Watts, Six degrees: The science of a Connected Age, W. W. Norton, 2004
- M. E. J. Newman, The structure and function of complex networks, SIAM Review 45, 167-256 (2003)
- Alderson D., Catching the Network Science Bug: ..., Operations Research, Vol. 56, No. 5, Sep-Oct 2008, pp. 1047-1065

Here are some basic notions about graphs that are useful to understand the routines in Section 0.2.

Figure 1 illustrates a general **directed** graph. The nodes are functions from this toolbox. An edge points from function A to function B if *function A is called within function B*. For example, *strongConnComp* is used within *tarjan*. Notice, also that *strongConnComp* points to itself, i.e. *strongConnComp* contains a recursion. Stand-alone functions, that use no other function, are **single nodes** in the graph, such as *leafNodes*, *getEdges* and *graphDual*.

A **directed graph** is a graph in which the links have a direction. In the functions graph one function can call another, but the call is usually not reciprocated.

A **single node** is a node without any connections to other nodes. *graphDual* is an example of a single node in Figure 1.

A **self-loop** is an edge which starts and ends at the same node. (*strongConnComp*→*strongConnComp*) is an example of a self-loop.

Multiedges are two or more edges which have the same origin and destination pair of nodes. This can be useful in some graph representations. In the functions graph this is equivalent to some function being called twice inside another function.

Basic graph statistics are the **number of nodes** (n) and the **number of edges** (m). The functions graph has 26 nodes and 17 edges.

The **link density** is derived directly from the number of nodes and number of edges: it is the number of edges, divided by the maximum possible number of edges.

$$density = \frac{m}{n(n-1)/2} \tag{1}$$

For the functions graph, the link density is about 0.05.

The **average nodal degree** is the average number of links per node. This is calculated as $2m/n$ (every edge is counted twice towards the total sum of degrees).

$$average\ degree = \frac{2m}{n} \tag{2}$$

The functions graph has 1.3 links per function on average.

A graph S is a **subgraph** of graph G, if the set of nodes (and edges) of S is subset of the set of nodes (and edges) of graph G.

A **disconnected** graph is a graph in which there are two nodes between which there exists no path of edges. In the functions graph there is no path between *averageDegree* and *subgraph*. So the functions graph is disconnected. Disconnected graphs consist of multiple connected components. The largest connected component (in number of nodes) is usually called the **giant component**.

In the context of **directed graphs**, the notion of strong and weak connectivity is important. A **strongly connected graph** is a graph in which there is a path from every node to every other node, where paths respect link directionality. In Figure 1, for example, there is a path from *strongConnComp* to *tarjan*, but no path in reverse. Therefore, the component (*strongConnComp*,*tarjan*) is not strongly connected. If, however, link directionality is disregarded, this subgraph is certainly connected. A **weakly connected graph** or subgraph is a graph which is connected if considered as undirected, but not connected if link directionality is taken into account.

0.2 Routines

0.2.1 getNodes.m

Returns the list of nodes for varying graph representations.

```
% Returns the list of nodes for varying graph representation types
% Inputs: graph structure (matrix or cell or struct) and type of structure (string)
%        'type' can be: 'adj','edgelist','adjlist' (neighbor list),'inc' (incidence matrix)
% Note 1: only the edge list allows/returns non-consecutive node indexing
% Note 2: no build-in error check for graph structure
%
```

```
% Example representations of a directed triangle: 1->2->3->1
%           'adj' - [0 1 0; 0 0 1; 1 0 0]
%           'adjlist' - {1: [2], 2: [3], 3: [1]}
%           'edgelist' - [1 2; 2 3; 3 1] or [1 2 1; 2 3 1; 3 1 1] (1 is the edge weight)
%           'inc' - [-1  0  1
%                     1 -1  0
%                     0  1 -1]
%
% GB: last updated, Sep 18 2012
```

0.2.2 getEdges.m

Returns the list of edges for varying graph representations.

```
% Returns the list of edges for graph varying representation types
% Inputs: graph structure (matrix or cell or struct) and type of structure (string)
% Outputs: edge list, mx3 matrix, where the third column is edge weight
%
% Note 1: 'type' can be: 'adj','edgelist','adjlist' (neighbor list), 'inc' (incidence matrix)
% Note 2: symmetric edges will appear twice, also in undirected graphs, (i.e. [n1,n2] and [n2,n1])
% Other routines used: adj2edgeL.m, adjL2edgeL.m, inc2edgeL.m
%
% Example representations of a directed triangle: 1->2->3->1
%           'adj' - [0 1 0; 0 0 1; 1 0 0]
%           'adjlist' - {1: [2], 2: [3], 3: [1]}
%           'edgelist' - [1 2; 2 3; 3 1] or [1 2 1; 2 3 1; 3 1 1] (1 is the edge weight)
%           'inc' - [-1  0  1
%                     1 -1  0
%                     0  1 -1]
%
% GB: last updated, Sep 18 2012
```

0.2.3 numNodes.m

Number of vertices/nodes in the network.

```
% Returns the number of nodes, given an adjacency list, or adjacency matrix
% INPUTs: adjacency list: {i:j_1,j_2 ..} or adjacency matrix, ex: [0 1; 1 0]
% OUTPUTs: number of nodes, integer
%
% GB: last update Sep 19, 2012

function n = numNodes(adjL)

n = length(adjL);
```

0.2.4 numEdges.m

Number of edges/links in the network.

```
% Returns the total number of edges given the adjacency matrix
% INPUTs: adjacency matrix, nxn
% OUTPUTs: m - total number of edges/links
```

```
%
% Note: Valid for both directed and undirected, simple or general graph
% Other routines used: selfloops.m, issymmetric.m
% GB, last updated Sep 19, 2012
```

0.2.5 linkDensity.m

The density of links of the graph. $Density = \frac{m}{n(n-1)/2}$ (n is the number of nodes and m is the number of edges).

```
% Computes the link density of a graph, defined as the number of edges divided by
% number_of_nodes(number_of_nodes-1)/2 where the latter is the maximum possible number of edges.
%
% Inputs: adjacency matrix, nxn
% Outputs: link density, a float between 0 and 1
%
% Note: The graph has to be non-trivial (more than 1 node).
% Other routines used: numNodes.m, numEdges.m
% GB: last update Sep 19, 2012
```

0.2.6 selfLoops.m

Number of selfloops, i.e. nodes connected to themselves.

```
% Counts the number of self-loops in the graph
%
% INPUT: adjacency matrix, nxn
% OUTPUT: integer, number of self-loops
%
% Note: in the adjacency matrix representation loops appear as non-zeros on the diagonal
% GB: last updated, Sep 20 2012
```

0.2.7 multiEdges.m

An edge counts towards the multi-edge total if it shares origin and destination nodes with another edge.

```
% Counts the number of multiple edges in the graph
% Multiple edges here are defined as two or more edges that have the same origin and destination nodes.
% Note 1: This creates a natural difference in counting for undirected and directed graphs.
%
% INPUT: adjacency matrix, nxn
% OUTPUT: integer, number of multiple edges
%
% Examples: multiEdges([0 2; 2 0])=2, and multiEdges([0 0 1; 2 0 0; 0 1 0])=2
%
% Note 2: The definition of number of multi-arcs (node pairs that have multiple edges across them)
% would be: mA = length(find(adj>1)) (normalized by 2 depending on whether the graph is directed)
%
% GB: last updated, Sep 20 2012
```

0.2.8 averageDegree.m

The average degree (# links) across all nodes. Defined as: $\frac{2m}{n}$, where n is the number of nodes and m is the number of edges. Also, $linkDensity = \frac{averageDegree}{n-1}$.

```
% Computes the average degree of a node in a graph, defined as
% 2 times the number of edges divided by the number of nodes (every edge is counted in degrees twice).
%
% Inputs: adjacency matrix, nxn
% Outputs: float, the average degree, a number between 0 and max(sum(adj))
%
% Note: The average degree is related to the link density, namely:
%       link_density = ave_degree/(n-1), where n is the number of nodes
%
% Other routines used: numNodes.m, numEdges.m
% GB: last update, September 20, 2012
```

0.2.9 numConnComp.m

Calculating the number of connected components in the graph by using the algebraic connectivity.

```
% Calculate the number of connected components using the Laplacian eigenvalues
%           - counting the number of zeros
%
% INPUTS: adjacency matrix, nxn
% OUTPUTs: positive integer - number of connected components
%
% Other routines used: graph_spectrum.m
% GB: last updated: September 22, 2012
```

0.2.10 findConnComp.m

findConnCompI.m: Finds the connected component to which node "i" belongs to.

```
% Find the connected component to which node "i" belongs to
%
% INPUTS: adjacency matrix and index of the key node
% OUTPUTS: all node indices of the nodes in the same group
%          to which "i" belongs to (including "i")
%
% Note: Only works for undirected graphs.
% Other functions used: kneighbors.m
% GB: last updated, Sep 22 2012
```

findConnComp.m: Find the connected components in an undirected graph.

```
% Algorithm for finding connected components in a graph
% Note: Valid for undirected graphs only
%
% INPUTS: adj - adjacency matrix, nxn
% OUTPUTS: a list of the components comp{i}=[j1,j2,...jk]
%
% Other routines used: findConnCompI.m, degrees.m
% GB: last updated, September 22, 2012
```

0.2.11 giantComponent.m

The largest connected component in a graph. Returns the set of nodes in the largest component, as well as its adjacency matrix.

```
% Extract the giant component of a graph;
% The giant component is the largest connected component.
%
% INPUTS: adjacency matrix, nxn
% OUTPUTS: giant component matrix and node indices
%
% Other routines used: findConnComp.m, subgraph.m
% GB: last updated: September 22, 2012
```

0.2.12 tarjan.m [8][9]

tarjan.m: Returns the strongly connected components in a directed graph.

```
% Find the stronly connected components in a directed graph
% Source: Tarjan, R. E. (1972), "Depth-first search and linear graph algorithms",
%                                 SIAM Journal on Computing 1 (2): 146-160
% Wikipedia description: http://en.wikipedia.org/wiki/Tarjan's_strongly_connected_components_algorithm
%
% Input: graph, set of nodes and edges, in adjacency list format,
%        example: L{1}=[2], L{2}=[1] is a single (1,2) edge
% Outputs: set of strongly connected components, in cell array format
%
% Other routines used: strongConnComp.m
% GB: last updated, Sep 22, 2012
```

strongConnComp.m: Support function for tarjan.m.

```
% Support function for tarjan.m
% "Performs a single depth-first search of the graph, finding all
% successors from the node vi, and reporting all strongly connected
% components of that subgraph."
% See: http://en.wikipedia.org/wiki/Tarjan's_strongly_connected_components_algorithm
%
% INPUTs: start node, vi;
%         graph structure (list), L
%         tarjan.m variables to update: S, ind, v, GSCC
% OUTPUTs: updated tarjan.m variables: S, ind, v, GSCC
%
% Note: Contains recursion.
% GB: last updated, Sep 22 2012
```

0.2.13 graphComplement.m

A graph with the same nodes, but "flipped" edges: where the original graph has an edge, the complement graph doesn't, and where the original graph doesn't have an edge, the complement graph does.

```
% Returns the complement of a graph
% The complement graph has the same nodes, but edges where the original graph doesn't and vice versa.
%
```

```
% INPUTs: adj - original graph adjacency matrix, nxn
% OUTPUTs: complement graph adjacency matrix, nxn
%
% Note: Assumes no multiple edges
% GB: last updated, September 23, 2012
```

0.2.14 graphDual.m

The graph dual is the inverted nodes-edges graph.

```
% Finds the dual of a graph; a dual is the inverted nodes-edges graph
% This is also called the line graph, adjoint graph or the edges adjacency
%
% INPUTs: adjacency (neighbor) list representation of the graph (see adj2adjL.m)
% OUTPUTs: adj (neighbor) list of the corresponding dual graph and cell array of edges
%
% Note: This routine only works for undirected, simple graphs.
% GB: last updated, Sep 23, 2012
```

0.2.15 subgraph.m

```
% This function outputs the adjacency matrix of a subgraph
%      given the supergraph and the node set of the subgraph.
%
% INPUTs: adj - supergraph adjacency matrix (nxn), S - vector of subgraph node indices
% OUTPUTs: adj_sub - adjacency matrix of the subgraph (length(S) x length(S))
%
% GB: last update, September 23, 2012
```

0.2.16 leafNodes.m

Leaf nodes are nodes connected to only one other node.

```
% Return the indices of the leaf nodes of the graph, i.e. all nodes of degree 1
%
% Note 1: For a directed graph, leaf nodes are those with a single incoming edge
% Note 2: There could be other definitions of leaves, for example: farthest away from a root node
% Note 3: Nodes with self-loops are not considered leaf nodes.
%
% Input: adjacency matrix, nxn
% Output: indices of leaf nodes
%
% GB: last updated, Sep 23, 2012
```

0.2.17 leafEdges.m

Leaf edges are edges with only one adjacent edge.

```
% Returns the leaf edges of the graph: edges with one adjacent edge only.
%
% Note 1: For a directed graph, leaf edges are those that "flow into" a leaf node.
% Note 2: There could be other definitions of leaves, for example: farthest away from a root node.
% Note 3: Edges that are self-loops are not considered leaf edges.
```

```
% Note 4: Single floating disconnected edges are not considered leaf edges.
%
% Input: adjacency matrix, nxn
% Output: set of leaf edges: a (num edges x 2) matrix
%         where every row contains the leaf edge nodal indices
%
% GB: last updated, Sep 23, 2012
```

References

[1] Milgram's small world experiment, source: http://en.wikipedia.org/wiki/Small_world_experiment, last accessed: Sep 23, 2012

[2] D.J. de S. Price, Networks of scientific papers, Science, 149, 1965

[3] D. Watts and S. Strogatz, Collective dynamics of 'small-world' networks, Nature 393, 1998

[4] S. Wasserman and K. Faust, Social network analysis, Cambridge University Press, 1994

[5] Duncan J. Watts, Six degrees: The science of a Connected Age, W. W. Norton, 2004

[6] M. E. J. Newman, The structure and function of complex networks, SIAM Review 45, 167-256 (2003)

[7] Alderson D., Catching the Network Science Bug: ..., Operations Research, Vol. 56, No. 5, Sep-Oct 2008, pp. 1047-1065

[8] Tarjan, R. E. , Depth-first search and linear graph algorithms, SIAM Journal on Computing 1 (2): 146-160, 1972

[9] Wikipedia description of Tarjan's algorithm, source: http://en.wikipedia.org/wiki/Tarjan's_strongly_connected_components_algorithm, last accessed: Sep 23, 2012