

Examination in School of Mathematical Sciences  
Semester 2, 2017

105637	APP MTH 4052	Applied Topic F: Complex Network ...
105661	APP MTH 7088	Applied Topic F: Complex Network ...
108646	STATS 4008	Stats Topic D: Complex Network ...
004013	STATS 7008	Stats Topic D: Complex Network ...

Official Reading Time: 10 mins  
Writing Time: 180 mins  
Total Duration: 190 mins

**NUMBER OF QUESTIONS: 6      TOTAL MARKS: 50**

**Instructions**

- Attempt all questions.
- Begin each answer on a new page.
- Examination materials must not be removed from the examination room.

**Materials**

- 1 Blue book is provided.
- Calculators without remote communications facilities are permitted.
- Students are permitted to bring two, double-sided pages of handwritten notes.
- English and foreign-language dictionaries may be used.

**DO NOT COMMENCE WRITING UNTIL INSTRUCTED TO DO SO.**

1. Consider the undirected, weighted graph shown in Figure 1.

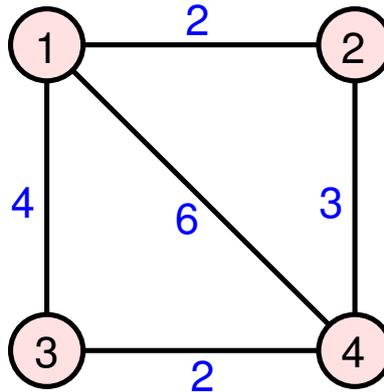


Figure 1: An undirected graph with link weights.

- (a) (i) Write the adjacency matrix for the graph.  
 (ii) From this, calculate the node degrees.
- (b) (i) State the handshake lemma relating the number of edges to the node degrees.  
 (ii) Prove it.  
 (iii) Illustrate the result for this graph.
- (c) (i) A local measure of clustering for each node  $i$  is defined by

$$c_i = \frac{|\{(j, k) \in E \mid j, k \in N_i\}|}{k_i(k_i - 1)/2},$$

where  $E$  is the set of edges in the graph,  $N_i$  is the neighbourhood of  $i$ , and  $k_i = |N_i|$ . Calculate this for each node, and hence calculate the average clustering coefficient for this network.

- (ii) Now compare this average to the *global* clustering coefficient, based on triangles and triples.
- (d) Calculate the shortest-paths between all pairs of nodes (using the distances defined on the graph, not hop counts) and from these determine
- (i) the *average distance* between nodes,  
 (ii) the *diameter* of the graph.

[14 marks]

2. (a) Construct the following graph products from the two graphs shown in Figure 2, and in each case, illustrate the result with a drawing of the graph.
- (i) The Cartesian product,
  - (ii) the Tensor product, and
  - (iii) the Rooted product (where the root node is indicated by the double circle).

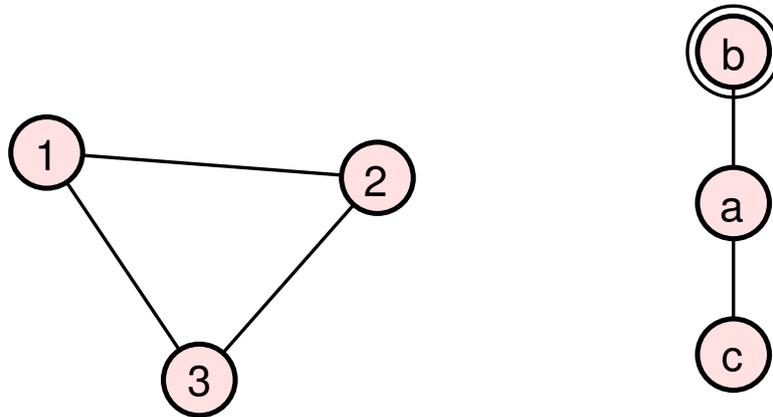


Figure 2: Two graphs. The root node is indicated by a double circle.

- (b) Describe the sense in which the Cartesian product operator on graphs commutes.

[7 marks]

3. Imagine an undirected random network with two different types of nodes. Nodes are connected independently with probability  $p_{ij}$ , which depends on the types of the two endpoints in the following manner

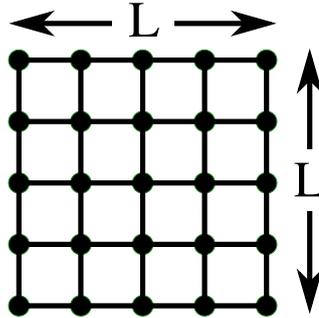
$$p_{ij} = \sqrt{q_{t(i)}q_{t(j)}},$$

where  $t(i) \in \{1, 2\}$  is the type of node  $i$ , and the  $q_k \in (0, 1)$  for  $k \in \{1, 2\}$  are values associated with nodes of each given type.

- (a) Assuming  $n_1$  and  $n_2$  nodes of each type, calculate the average degree of
- Type 1 nodes,
  - Type 2 nodes, and
  - a random node.
- (b) Calculate the asymptotic clustering coefficient for this network as the number of nodes grows. Assume that the graph is sparse in the sense that as the graph grows, the average degree of each type of node remains (approximately) constant.

[10 marks]

4. (a) A network consists of a single clique of  $n \geq 3$  nodes. Does it contain an Eulerian cycle, when
- (i)  $n$  is even?
  - (ii)  $n$  is odd?
- (b) Imagine a  $d$ -dimensional hyper-cube formed from square lattices with  $L$  edges (and  $L + 1$  vertices) along each side, as shown below in 2-dimensions.



- (i) What is the diameter of the graph as a function of the dimension  $d$ ?
- (ii) What is the diameter of the graph as a function of the total number of nodes  $n$ ?

[6 marks]

5. The following table presents a set of “reads” of a genome in lexicographic order. The multiplicity of each read is how often it occurred.

Assemble the reads into a complete genome. List all possible sequences that could have produced these reads.

read	multiplicity
AAC	1
ACT	1
CTG	2
GAA	1
GCT	1
GTA	1
TGA	1
TGC	1
TGT	1

[6 marks]

6. The paper “How to infer gene networks from expression profiles,” Bansal *et al.*, *Molecular Systems Biology*, Vol.3, 2007, describes and compares methods for reconstruction gene-gene expression networks, *i.e.*, the networks of genes that are commonly expressed together.

The authors aim, in this paper, to test alternative algorithms for reconstruction using simulated networks where we know the true network, and can thus exactly measure the reconstruction error. The simulation model they use is a “random” network, *i.e.*, the Gilbert-Erdős-Rényi model. They test performance of algorithms on networks of various sizes and average node degree using this model.

The “The Network Data Repository,” presents a large set of human gene-gene interaction network data at <http://networkrepository.com/bio-human-gene1.php>. The following table lists some of the metrics of the network described in this data.

number of nodes	21.9 K
number of edges	12.3 M
maximum degree	7.9 K
average degree	1.1 K
assortativity	0.136
clustering coefficient	0.621

Using these network metrics, comment on the validity of the simulations of Bansal *et al.*

You should provide a concise, but well-argued paragraph or two answering the question. Marks will be based on the presentation of the argument as well as the content. Marks will be deducted if your argument is diluted by extraneous or irrelevant points.

**[Note: neither the paper or data are attached – you don’t need to see these details to answer the question.]**

[7 marks]