SMM638 — Mid-Term Project

Project Package

Submission deadline	November 21st at 4:00 PM
Template	Mandatory, available as .docx and .tex format
Wordl imit	True, see template documents

© Simone Santoni

Contents

1 Project Description

1.1 Silico Inc., Company Background

For this MTP, you will work with real-world data on the R&D function of a large company in the industry of semiconductors. Per the non-disclosure-agreement that is in place, i) I refer to the company with the fictionary name of Silico Inc., ii) the data do not contain any personal information (e.g., an engineer's name) and, iii) the distribution of the gender variable was contaminated by using synthetic data generation techniques (see Section 2).

The R&D function of Silico Inc. has 19 units — each of which is specialized in a field such as microelectromechanical systems, ultra-low-voltage processors, and rendering — and includes $1,158$ engineers. [1]

The technological innovation process of Silico Inc. is based on medium-scale projects whose duration ranges between a few months and one year and a half. The KPI of a project is the creation of patentable technology. Put simply, the projects generating a patent are valuable projects for Silico Inc.; less so, the projects who fail to create a robust and original technology (hence they do not lead to any patents!).

In terms of organizing, unit heads form small teams of engineers[2] *ad hoc* to work on specific technical issues. At the end of the project, the teams dissolve and the individual engineers join new projects.

In terms of process, all Silico Inc. projects follow the same roadmap:

- In the first phase, the head of the unit communicates the project's technical requirements to the selected group of engineers (e.g., reducing the size of a biosensor[3] currently[4]sold by Silico Inc. of 20%);

- In the second phase, the engineers create a five to ten-page project proposal;

- In the third phase, the head of a unit i) evaluates the project proposal against several criteria, ii) provides the team with a structured feedback, and iii) decides the project budget;

---

[1]All throughout the various sections of the document, I will use the term engineer and employee interchangeably.

[2]The modal team size is five members. Rarely, teams comprise six or seven members.

[3]For basic information on biosensors, see `https://en.wikipedia.org/wiki/Biosensor`

[4]The modal team size is five members. Rarely, teams comprise six or seven members.

- In the fourth phase, the engineers try to develop a technical to the problem identified in the first phase. If the solution is robust and original, a patent application is filed for consideration by the United States Patent and Trademark Office (USPTO).[5]

1.2 Expectations: Your Role

Make the following assumptions. First, you are a junior business analyst working for a global consultancy company. Second, Silico Inc. is your employer's client. Third, you are supposed to contribute to creating a deck to use for a meeting with the client. Specifically, you are asked to create a set of network analytic models helping the client to improve the performance of the R&D projects.

1.3 Objective

Your network analytic models should provide actionable insights on how to improve the performance of R&D projects by assembling the right team of engineers.

---

[5]For your reference, the USPTO provides an overview of the patenting process at `https://www.uspto.gov/patents/basics/patent-process-overview`.

2   Data

The client provided you with data on the engineers affiliated with the 19 R&D units, the composition and performance of a sample of past projects, and the network of information exchange between engineers. The data span four files:

A. `ua.graphml`, reporting the affiliations of employees with the units of the R&D function of Silico Inc.;

B. `pa.graphml`, reporting the affiliations of employees with a sample of projects carried out within the R&D function of Silico Inc. recently;

C. `po.csv`, reporting the performance for a sample of past projects;

D. `ie.graphml`, reporting the information exchange relationship among the employees of Silico Inc. in the R&D function in the sample.

Table 1 illustrates the key features of the network data files mentioned in points A, B, and D. These data are stored in GraphML, a format based on XML and hence ideally suited as a *"common denominator for all kinds of services generating, archiving, or processing graphs."*[6] NetworkX has extensive capabilities to read and write network data in many formats, including GraphML files. The students are encouraged to refer to the section of NetworkX's references `Reading and Writing Graphs > GraphML`.[7]

---

[6]See `http://graphml.graphdrawing.org/`
[7]See `https://networkx.org/.../reference/readwrite/graphml.html`.

Table 1: Overview of the Network Data

Data file	Relationship	Network Form	Node Attributes	Edge Attributes
ua.graphml	Employee-unit affiliation	Two-mode	●	○
pa.graphml	Employee-project affiliation	Two-mode	●	○
ie.graphml	Employee-employee information exchange	One-mode, undirected	●	●

Notes. — ● denotes that the attribute is present in the data file; ○ denotes that the attribute is not present in the data file.

```
✓  ua = nx.read_graphml("ua.graphml") ⋯

✓  nx.is_bipartite(ua) ⋯
⋯  True


✓  list(ua.nodes(data=True))[0] ⋯
⋯  ('11-1', {'bipartite': 0})


✓  list(ua.edges(data=True))[0] ⋯
⋯  ('11-1', '11', {})
```

Figure 1: A preview of the attributes of the `ua.graphml` network dataset

2.1 Employee-Unit Affiliation Network

Per Figure 1, `ua.graphml` is a two-mode network connecting 1,158 employees to the 19 units in the R&D function of Silico Inc. Employee nodes' labels combine the ID of the unit of affiliation and the ID of the employee (e.g., '11-1'); unit nodes' labels reflect the ID of the units (e.g., '11'). The nodes have one attribute, `bipartite`, discriminating between employees (`bipartite = 0`), and units (`bipartite = 1`).

2.2 Employee-Project Affiliation Network

Per Figure 2, `pa.graphml` is a two-mode network connecting a 665 employees[8] to the sample of 133 past projects. Project nodes' labels combine the ID of the unit of affiliation[9] and the ID of the project (e.g., '11-p5'). The nodes have one attribute, `bipartite`, discriminating between employees (`bipartite = 0`), and projects (`bipartite = 1`).

---

[8]It is self-evident from the data that there are 493 employees who are not affiliated with any projects in the sample. The reason is that, at the time of the data collection, these employees were working on ongoing projects for which no outcome was available still.

[9]In the interest of redundancy, let me stress that the individual projects are owned by the units of the R&D function of Silico Inc.

```
✓  pa = nx.read_graphml("pa.graphml") ···


✓  nx.is_bipartite(pa) ···
··· True


✓  list(pa.nodes(data=True))[0] ···
··· ('11-1', {'bipartite': 0})


✓  list(pa.edges(data=True))[0] ···
··· ('11-1', '11-p5', {})
```

Figure 2: A preview of the `pa.graphml` network dataset

2.3   Information Exchange Network

The dataset `ie.graphml` is a one-mode network connecting the 1,158 employees via the information exchange relationship. The data come from a survey administered to the population of employees, who were asked two things:

- To indicate the names of colleagues with whom they shared information on subjects such as technological trends, technical challenges and possible solutions;

- To indicate the strength of each tie with the respect to the frequency and intensity of the information exchange.

Per Figure 3, each node has three attributes: `gender` (0 = female employee; 1 = male employee), `ti_exp`, the technological innovation experience of employees, measured by the counts of granted patents in which an engineer 'inventor' or 'co-inventor', and `tenure`, the years elapsed since the employee joined Silico Inc.[10]; Even the edges have an attribute in this network: `strenght` is equal to 0 for 'weak ties' — that is, ties with limited frequency and intensity — and equal to 1 for 'strong ties' — that is, ties with

---

[10]For some employees hired in the vicinity of the data collection, `tenure` is equal to 0.

```
✓ ie = nx.read_graphml("ie.graphml") ⋯


✓ list(ie.nodes(data=True))[0] ⋯
⋯   ('11-1', {'gender': 1, 'ti_exp': 8.0, 'tenure': 19.04362767048749})


✓ list(ie.edges(data=True))[0] ⋯
⋯   ('11-1', '11-0', {'strength': 1})
```

Figure 3: A preview of the `ie.graphml` network dataset

substantial frequency and intensity. Figure 4 illustrates the topology of the `ie.graphml` network.

2.4 Project Performance

Table 2 indicates the performance measures for the 133 projects in the sample. The quality of a project proposal is assessed by the head of the unit in the third phase of the project's life cycle. The patent application is submitted by the unit after the fourth phase provided the project generated a robust and novel technological innovation.

Table 2: Project Performance Measures

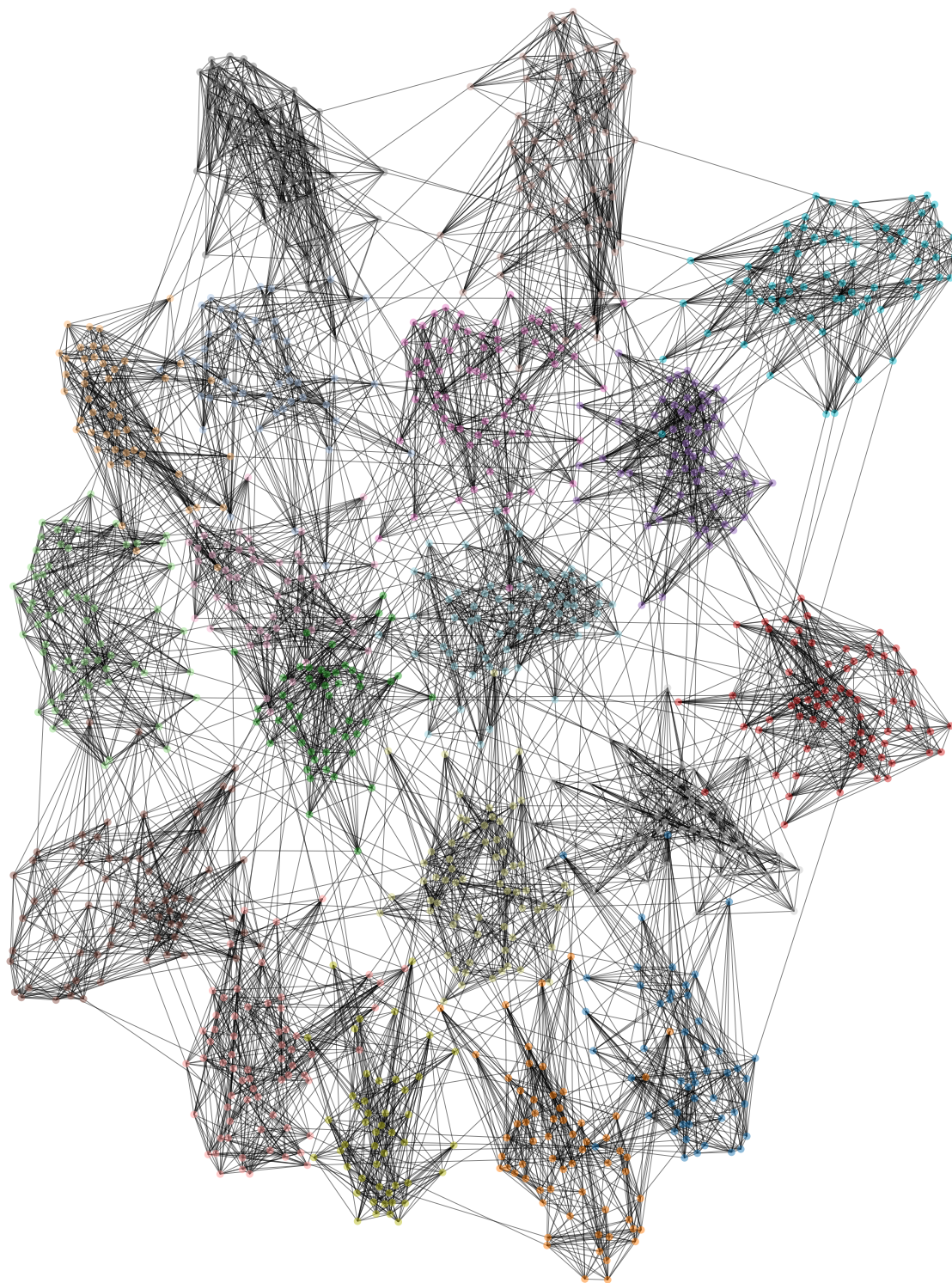| Measure	Synopsis	N
Project proposal quality score . .	The score assigned by the head of the unit to the project proposal (out of 100)	133
Patent application.	The project generated a patent application (0 = No, 1 = Yes)	133

9

Figure 4: A visualization of the ie.graphml network dataset. *Notes.* — the color of the nodes reflect the unit of affiliation of the employees; in the interest of simplicity, the strength of the ties is not reported in the visualization.

3   Deliverables

By November 21st, the students must submit:

- An executive summary covering the following aspects

    - The main steps of the proposed network analysis;

    - The justification for each main step;

    - The main results of the network analysis;

    - A set of actionable business analytics recommendations grounded in the network analysis results;

- The Python code necessary to replicate the network analysis results reported in the executive summary.

The students are required to use the provided template — available in .docx and .tex format — and respect the word limits applying to the various boxes/sections of the template. The total length of the document is 2,500 words.

4   References and Contextual Information

The readings administered in the third and fourth weeks of the module offer a solid knowledge platform to carry out this MTP. *Ditto*, the students are welcome to liaise with the module leader about further readings and contextual information on the role of networks in the management of innovation.