Graphs, II

李昂生

Discrete Mathematics
U CAS
June, 2018

Outline

1. Representations
2. Approaches
3. Fundamental question
4. Connectivity
5. Small world phenomenon
6. Cut
7. Invariants
8. Exercises

General view

- Fundamental problems
- Basic methods
  - combinatorial methods
  - probabilistic method
  - algebraic method

Figures and Tables

- Figures
  - intuitive
  - it is hard to represent large graphs
- Tables
  - clear
  - hard to express large graphs
- None of these representations support mathematical study of graphs

Adjacency Matrices

As mentioned before. For completeness, we describe as follows.
Let $G = (V, E)$ be a simple graph with $|V| = n$. Fix an ordering of the vertices of G by

$$v_1, v_2, \cdots, v_n. \tag{1}$$

The adjacency matrix （邻接矩阵）A (or A_G) of G is the $n \times n$ matrix $A = (a_{ij})$, where

$$a_{ij} = \begin{cases} 1, & \text{if } \{v_i, v_j\} \in E, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Key: Graphs have been realised by algebra.

Incidence Matrices

Let $G = (V, E)$ be an undirected graph. Suppose that

(a) Fix an ordering of the vertices as follows:

$$v_1, v_2, \cdots, v_n.$$

(b) Fix an ordering of the edges as follows:

$$e_1, e_2, \cdots, e_m.$$

Then incidence matrix （连接矩阵）$M = (m_{ij})$ is defined as follows:

$$m_{ij} = \begin{cases} 1, & \text{if } v_i \text{ is incident with } e_j, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Combinatorial Approach to Graphs

This is the mainstream of classical graph theory.

Spectral Graph Theory

- Matrix theory and linear algebra were used to study graphs **The successful**: to study the graphs that are regular, symmetric.
- The key is to explore the role of eigenvalues
- Early applications are in Chemistry; eigenvalues were associated with the stability of molecules
- More applications are in theoretical physics and quantum mechanics, for example, in minimising the energies of Hamitonian systems
- Current applications: expanders, communication networks, algorithms

Question Is there a geometrical study of graphs? Yes, there is an interaction between spectral graph theory and differential geometry.
This is a new direction, calling for more and better investigations.

Information Theoretical Approach to Graphs

- Using Shannon's entropy to measure the structural complexity of chemical graphs
- Using Shannon's metric to study physical systems, including both classical and quantum systems
- Structural information, proposed by myself and Pan, 2016:
  – to study the complexity of interactions, communications and operations in graphs
  – to distinguish the order from disorder in graphs
  – to determine and decode the natural or true structure of structured noisy data

Graph Algorithms

Graph algorithms form the mainstream of algorithms.

Graph Isomorphism Problem

Definition 1

Given two simple graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, we say that G_1 is **isomorphic to** G_2, if there is a one-to-one map f from V_1 to V_2 such that for any $x, y \in V_1$,

$$(x, y) \in E_1 \iff (f(x), f(y)) \in E_2. \tag{4}$$

In this case, we write $G_1 \cong G_2$ via f.

Graph isomorphism problem (GI): To decide for two graphs G, H, whether or not $G \cong H$.

Open Question: If there is polynomial time algorithm for GI?

Significance of GI

- The relation $\cong$ $(=)$ is fundamental to all math
- Molecular graphs, chemical compounds
- Electronic circuits
- Rich source for algorithmic ideas
  WHY?

Similarity

Since $\cong$ is hard, real graphs are large, and single vertex or edge may not be so essential, we may relax the relation $\cong$ to *similarity*.

This leads to the question to measure the similarity of graphs.

There are many definitions of similarity in different applications.

Interactive Proof (IP) - An Algorithmic Idea from NGI

Let GI be the set of all pairs (G, H) such that $G \cong H$, and let NGI be the set of all pairs (G, H) such that $G \ncong H$ where G, H are graphs.

Definition of Interactive Proof (IP)

An **interactive proof system** is a pair $\langle P, V \rangle$ of *prover* P and *verifier* V such that

- The prover P is an algorithm with unlimited power
- The verifier V is a probabilistic polynomial time algorithm
- The verifier V decides to either accept or reject an input instance x based on its own computation and the interactions with the prover P.

Examples

1. A reviewer decides to accept or reject a paper for publication
   Now the prover is the paper written by the authors of the paper, the verifier is the referee.
2. Tea test

Interactive Proof System for NGI

The interactive proof system $\langle P, V \rangle$ proceeds as follows: Given two graphs G_1 and G_2,

(1) - V picks $i = 1$ or 2 randomly,
- V picks a random permutation π,
- set $H = \pi(G_i)$, and
- V sends H to P, asking for the index i, but V keeps i as a private key.

(2) P chooses an index j and send it to V

(3) V accepts if $i = j$, and rejects otherwise.

Completeness

Assume that $G_1 \ncong G_2$.
In this case, the honest and powerful prover P certainly knows the index i from the graph H provided by the verifier V.
Therefore, by interacting with this prover P, verifier V accepts with probability 1.
This means that there is a prover P such that by communicating with P, the verifier V accepts the input instance (G_1, G_2) with probability 1.

Soundness

Assume $G_1 \cong G_2$.
In this case, for any prover P, from the random graph H received from the verifier V, whatever the index j is chosen by the prover P, the probability that $i = j$ is $\frac{1}{2}$.
Therefore, for any prover P, the probability that V accepts through the interaction with P is $\frac{1}{2}$.

The prover P

The prover P could be an arbitrary function or an algorithm with unlimited power.

The Verifier V

- The verifier V is a **probabilistic polynomial time algorithm**.
- The interactive proof system for NGI is the first such system in history. It leads to a completely new direction in computational complexity.
- In fact, this leads to the modern approach to complexity theory.

Interactive Proof System and Zero-Knowledge

The interactive proof system for NGI also leads to the modern approach to cryptography.

The new theory is the study of **zero-knowledge protocols** with the intuition as mentioned in the example before.

Paths

Definition 2

Let $G = (V, E)$ be an undirected graph and n be a nonnegative integer.
A *path* of length n from u to v in G is a sequence of vertices of the form:

$$u = x_0, x_1, \cdots, x_{n-1}, x_n = v,$$

such that for each $i < n$, $(x_i, x_{i+1}) \in E$.
If $u = v$, the path is called a *cycle* or *circuit*.
A path is also referred to as a *walk*.

The same notions can be similarly defined for directed graphs.

Connectivity

Definition 3

Let $G = (V, E)$ be an undirected graph. We say that G is *connected*, if for any two vertices u and v, there is a path from u to v in G.

An undirected graph that is not connected is called *disconnected*.

Connected Components

Definition 4

Let $G = (V, E)$ be an undirected graph. A *connected component* of G is a maximal connected connected subgraph of G.

Theorem 5

If G is disconnected, then G is the disjoint union of all its connected components.

Distance and Diameter

Definition 6

Let $G = (V, E)$ be an undirected and connected graph and u, v be two vertices. The *distance of u and v in G* is the length of the shortest path from u to v in G.
We use $d(u, v)$ to denote the distance of u and v in G.
The *diameter of G* is defined as

$$D(G) = \max\{d(u, v) \mid u, v \in V\}. \quad (5)$$

Let $G = (V, E)$ be an undirected graph. If G is disconnected, then the *diameter of G* is the maximal diameter of all its connected components of G.

Shortest Path

Recall:
Let $G = (V, E)$ be an undirected and connected graph.
For any $u, v \in V$, the distance of u and v in G is at most n, where $n = |V|$.
One word proof: Let P be a shortest path from u to v in G. Then for any vertex $x \in V$, x appears in P at most once, unless $x = u$ or v, for which the path is a cycle.
Why we need paths?

- People walk everyday, following a shortest path.
- People easily find a short path for their global travel.
- Animals know their shortest paths
- The world is a highly connected small world.

Social Science Experiments

In the 1960's, Stanley Milgram implemented a social science experiment:

- Any two individuals were connected by a short sequence of acquaintances.
- Ask a source individual to send a letter to some one who may be close to the target individual. The basic information including the address and occupation are given to each sender. In successful experiments, it would take on average 5 to 6 steps for a letter to reach its target.
  Key: the letter was asked to send someone who may be close to the target.
- The results was referred to as the so called "six degree of separation".

Small World Phenomenon

Given a network $G = (V, E)$ of n vertices for large n, we say that G has the *small world phenomenon*, if:

1. (Small diameter property) The diameter of G is $O(\log n)$ or $\log^{O(1)} n$ where $n = |V|$.
2. (Clustering effect) Two vertices sharing common neighbours are more likely to be linked.

Many real-world networks have the small world phenomenon.

Algorithmic Small World Phenomenon

For a network with small world phenomenon, the algorithm finds a path from one vertex to another in time $\mathrm{poly}(\log n)$.

Open Question: Characterise the networks have the algorithmic small world phenomenon.

Small World Model

Kleinberg, 1998:
Consider a two-dimensional $n \times n$ grid where each vertex is connected to its four adjacent vertices, referred to as *local edges*. In addition to these local edges, there is one or q for some constant q long distance edge out of each vertex.
For every vertex u, the probability that there is a long distance edge from u to v is proportional to

$$\frac{1}{d^r(u,v)}, \tag{6}$$

for some constant $r \geq 0$, $d(u,v)$ is the distance between u and v using only local edges.

Normalising Constant

- The probability that the long distance edge from u to v is proportional to $\frac{1}{d^r(u,v)}$.
- For $k \leq \frac{n}{2}$, the number of vertices at distance exactly k from u is at most $4k$.
- The normalising constant is defined by

$$c_r(u) = \sum_v d^{-r}(u,v). \tag{7}$$

$r > 2$

(1)

$$\begin{aligned} c_r(u) &= \sum_v d^{-r}(u,v) \\ &\geq \sum_{k=1}^{\frac{n}{2}} k \cdot k^{-r} \\ &= \sum_{k=1}^{\frac{n}{2}} k^{1-r} \\ &= c \geq 1. \end{aligned}$$

$r > 2$ - continued

By (1), we have
(2) The probability of an edge from u to v is

$$\frac{d^{-r}(u,v)}{c_r(u)} \leq d^{-r}(u,v).$$

According to (2):
(3) The probability that a long edge is of length $\geq k = n^{\frac{r+2}{2r}}$ is upper bounded by

$$(n^{\frac{r+2}{2r}})^{-r} = n^{-\frac{r+2}{2}}.$$

$r > 2$ - continued

(4) By linearity of expectation, the expected number of edges of length at least $n^{\frac{r+2}{2r}}$ is at most $n^2 \cdot n^{-\frac{r+2}{2}} = n^{\frac{2-r}{2}}$, which goes to 0 as n goes to infinity.

$r > 2$ - continued

(5) For at least one half of the pairs of vertices (the grid distance measured by grid edges) between the vertices, is greater than or equal to $\frac{n}{4}$. Any path between them must have at least

$$\frac{n/4}{n^{\frac{r+2}{2r}}} = \frac{1}{4} \cdot n^{\frac{r-2}{2r}}$$

edges, since there are no edges longer than $n^{\frac{r+2}{2r}}$, and so there is no polylog length path.

$r = 2$

The normalising constant $c_r(u)$ is upper bounded by

$$
\begin{aligned}
c_r(u) &= \sum_v d^{-r}(u, v) \\
&\leq \sum_{k=1}^{2n} (4k) \cdot k^{-2} \\
&= 4 \sum_{k=1}^{2n} \frac{1}{k} \\
&= \Theta(\ln n).
\end{aligned}
$$

$r = 2$- continued

For u, v, suppose that $d(u, v) = k$. Let $B(v, \frac{k}{2})$ be the set of vertices x such that $d(v, x) \leq \frac{k}{2}$. Then

$$|B(v, \frac{k}{2})| = \Theta(k^2).$$

Therefore, the probability that there is a long distance edge from u to a vertex in the ball $B(v, \frac{k}{2})$

$$\Theta(k^2 \cdot k^{-2} \cdot \frac{1}{\ln n}) = \Theta(\frac{1}{\ln n}).$$

$r = 2$ - continued

Consider $\Omega(\ln^2 n)$ steps of the path from u. The long-distance edges from the points visited at these steps are chosen independently and each has probability $\Omega(\frac{1}{\ln n})$ of reaching within $\frac{k}{2}$ of v.
The probability that none of them does is

$$(1 - \Omega(\frac{1}{\ln n}))^{c \ln^2 n} = c_1 e^{-\ln n} = \frac{c_1}{n},$$

for an appropriate choice of constants c and c_1.

$r = 2$ - continued

This shows that
– almost surely, after $\Omega(\ln^2 n)$ many steps random walk from u, we arrive at a vertex in the ball $B(v, \frac{k}{2})$.
This further implies that
after $O(\ln^3 n)$ steps random walk, with high probability, we have visited the vertex v.
This is a randomized algorithm that finds a path from u to v in time $O(\ln^3 n)$.

$r < 2$

The normalising constant $c_r(u)$ is

$$
\begin{aligned}
c_r(u) &= \sum_v d^{-r}(u,v) \\
&\geq \sum_{k=1}^{n/2} k \cdot k^{-r} \\
&\geq \sum_{k=n/4}^{n/2} k^{1-r} \\
&= \Omega(n^{2-r}).
\end{aligned}
$$

$r < 2$- continued

In this case, there is no polylog time algorithm that finds a short path.
Let $\delta = \frac{2-r}{4}$.
We show that a local algorithm cannot find paths of length $O(n^\delta)$.
Suppose that the algorithm finds a path with at most n^δ edges from u to v. There must be a long-distance edge on the path which terminates within distance n^δ of v; otherwise, the path would end in n^δ grid edges and would be too long.
There are $O(n^{2\delta})$ vertices within distance n^δ of v.

$r < 2$ - continued

The probability that there is a long edge from u to a vertex in the ball $B(v, n^{\delta})$ is at most

$$n^{2\delta} \cdot \frac{1}{n^{2-r}} = n^{\frac{r-2}{2}}.$$

The probability that any vertex in a path of length n^{δ} is in ball $B(v, n^{\delta})$ is at most

$$n^{\frac{r-2}{2}} \cdot n^{\delta} = n^{\frac{r-2}{4}} = o(1).$$

The challenge

Kleinberg's negative results contradict Stanley Milgram's experiment.
How to solve this problem?
There is a new theory of algorithmic small world phenomenon, a project in progress

Cut

Suppose that $G = (V, E)$ is a connected computer network.

(1) A vertex $v \in V$ is called a *cut vertex*, if the graph G' obtained from G by deleting v and all incident edges is disconnected.

(2) An edge $e \in E$ is called a *cut edge* or *bridge*, if the graph G' obtained from G by deleting edge e is disconnected.

Question: Find the cut vertices and cut edges, if any.

Vertex Connectivity

Let $G = (V, E)$ be a connected graph and $V' \subset V$. We say that V' is a vertex cut of G, if $G - V'$ is disconnected.
We define the *vertex connectivity* of G is the minimum size of the vertex cuts of G.
We use $\kappa(G)$ to denote the vertex connectivity of G.
We say that G is *k-connected*, if $\kappa(G) \geq k$.

Edge Connectivity

Let $G = (V, E)$ be a connected graph. We define the *edge connectivity* of G to be the minimum size of cuts of G where a cut C of G is a subset of E whose removal disconnects G. We use $\lambda(G)$ to denote the edge connectivity of G.

Theorem 7

Let $G = (V, E)$ be a connected graph. Then

$$\kappa(G) \leq \lambda(G) \leq \min_{v \in V} d(v).$$

Application: ?

Connectivity in Directed Graphs

Let $G = (V, E)$ be a directed graph.

1. We say that G is *strongly connected*, if there is a path from a to b and from b to a whenever a and b are vertices of G.
2. We say that G is *weakly connected*, if the underlying undirected graph G' of G is connected.

Strong Components of a Directed Graph

Definition 8

Let $G = (V, E)$ be a directed graph. A *strongly connected component* or *strong component* of G is a maximal strongly connected subgraph of G.

Theorem 9

Let G be a directed graph. Then G is the disjoint union of all its strong components.

Giant Component

Let $G = (V, E)$ be a graph, directed or undirected. Let H be the largest strong component for directed G or component for undirected G.

We say that G has a *giant component*, if H contains a large fraction of the vertices of G.

Experimental Result: The web graph, and many real-world networks have a giant component.

Robustness Against Physical Attack

Physical attack means to delete vertices or edges of a graph.
Given a network G, suppose that G is connected.
The robustness requires that for any attack of a an appropriately large number of vertices or edges,

- the remaining graph has a giant component
- the diameter of the remaining graph is still very small.

Experimental result Most real-world networks are robust against physical attacks.
The art of war?

Invariants

If $G \cong H$, then

- the numbers of vertices and edges of G and H are the same
- the degree distributions of G and H are the same
- the structural properties of G and H are the same

Counting the Paths

Theorem 10

Let $G = (V, E)$ be a graph and $A = A_G$ be an adjacency matrix. Then the number of different paths of length r from v_i to v_j equals the $(i.j)$-entry of A^r.

Proof.

By induction on r. □

Exercises - 1

(1) Show that every connected graph with n vertices has at least $n-1$ edges.

(2) Show that in a connected simple graph, there is a path from every vertex of odd degree to some other vertex of odd degree.

(3) Suppose that v is an endpoint of a cut edge. Prove that v is a cut vertex if and only if this vertex is not pendant.

(4) Show that a vertex u in a connected simple graph G is a cut vertex if and only if there are vertices x and y, both different from u such that every path between x and y passes through u.

Exercises - 2

(5) Show that a simple graph with at least two vertices has at least two vertices that are not cut vertices.

(6) Show that an edge in a simple graph is a cut edge if and only if this edge is not part of any simple circuit in the graph.

(7) Show that a simple graph with n vertices and k connected components has at most $\frac{(n-k)(n-k+1)}{2}$ edges.

(8) Show that a simple graph G with n vertices is connected if it has more than $\frac{(n-1)(n-2)}{2}$ edges.

谢谢！