Trees

March 2, 2010

1 Forests and Trees

- A graph is said to be **acyclic** if it does not contain cycle. An acyclic graph is also called a **forest**. A connected acyclic graph is called a **tree**. A tree is said to be **trivial** if it contains a single vertex.
- Every nontrivial tree has at least two leaves, i.e., two vertices of degree 1.
- Any two vertices in a tree are connected by a unique path.
- If T is a tree, then $|E(T)| = |V(T)| - 1$. If F is a forest, then $|E(F)| = |V(F)| - c(F)$, where $c(F)$ is the number of connected components of F.

2 Rooted Trees

- A tree T with a specified vertex x is called a **rooted tree** with the **root** x, denoted $T(x)$. A tree with a root x is referred to an x**-tree**.
- A **branching** is a rooted tree with an orientation such that every vertex but the root has in-degree 1. A branching with a root x is referred to an x**-branching**.
- Let D be a digraph and x a vertex of D. Let X be the set of vertices reachable from x in D. Then there exists an x-branching $T(x)$ in D with $V(T) = X$.

3 Spanning Trees

- A **spanning subgraph** H of a graph G is a subgraph such that $V(H) = V(G)$. A **spanning tree** T of a connected graph G is a tree and is also a spanning subgraph of G.
- Every connected graph has a spanning tree.
- If T is a spanning tree of a graph G, then for any edge $e \in E(G) - E(T)$, there exists a unique cycle C_e in $T \cup e$. Such a cycle is called a **fundamental cycle** of G with respect to T.

  Proof. Let u, v be end-vertices of e. There are two internal-vertex disjoint paths between u and v in $T \cup e$, i.e., uev and the unique path between u and v in T. The two paths form a cycle C_e in $T \cup e$.

  Suppose C'_e is a cycle other than C_e in $T \cup e$. Then $C_e \Delta C'_e$ is a nontrivial even graph contained T. This is a contradiction for T contains no cycle. □

- If T is a spanning tree of a graph G, then for any edge $e \in E(T)$, there exists a unique bond B_e of G which is contained in $T^c \cup e$. Such a bond is called a **fundamental bond** of G with respect to T.

Proof. Remove the edge e from T to obtain two disjoint trees T_1 and T_2. Then the edge set between the vertex sets $V(T_1)$ and $V(T_2)$ is a bond of G.

Suppose B'_e is a bond of G other than B_e and is contained in $T^c \cup e$. Then the symmetric difference $B_e \Delta B'_e$ is an edge cut of G and is contained in T^c. Clearly, $E(T)$ is contained in $E(G) - B_e \Delta B'_e$. Since T is connected, so is $E(G) - B_e \Delta B'_e$. This a contradict to that $B_e \Delta B'_{\boldsymbol{e}}$ is an edge cut of G. ☐

Theorem 3.1. *The collection of fundamental cycles of a connected graph G with respect to a spanning tree T forms basis of the cycle space of G. So the cycle space has dimension $|E(G)| - |V(G)| + 1$.*

Proof. Let $\{C_e \mid e \in E(G-T)\}$ be the collection of fundamental cycles of G with respect to T. We first show that $\{1_{C_e} \mid e \in E(G-T)\}$ is linearly independent over $\mathbb{F}_2$.

Assume $\sum_{e \in E(G-T)} a_e 1_{C_e} = 0$ for some coefficients a_e in $\mathbb{F}_2$. For a particular edge $e_0 \in E(G-T)$, we have $e_0 \in C_{e_0}$ and $e_0 \notin C_e$ for all $e \in E(G-T)$ such that $e \neq e_0$. So the left-hand side of $\sum_{e \in E(G-T)} a_e 1_{C_e} = 0$ is a_{e_0}. Thus $a_{e_0} = 0$. This proves the linear independence.

To show that $\{1_{C_e} \mid e \in E(G-T)\}$ spans the cycle space of G, given a cycle C of G. We claim that $1_C = \sum_{e \in E(C-T)} 1_{C_e}$, i.e.,

$$1_C + \sum_{e \in E(C \cap T^c)} 1_{C_e} = 0. \tag{1}$$

It is clear that the left-hand side of (1) is zero on all edges of $E(T^c - C)$. The left-hand side of (1) also cancels to zero on the edges of $E(C \cap T^c)$. Thus the left-hand side of (1) is the characteristic function of an even graph on T. Since T does not contain cycle, it follows that the left-hand side of (1) is zero on all edges of G. ☐

Theorem 3.2. *The collection of fundamental bonds of a connected graph G with respect to a spanning tree T forms basis of the bond space of G. So the bond space has dimension $|V(G)| - 1$.*

Proof. Let $\{B_e \mid e \in E(T)\}$ be the collection of fundamental bonds of G with respect to T. We first show that $\{1_{B_e} \mid e \in E(T)\}$ is linearly independent over $\mathbb{F}_2$.

Assume $\sum_{e \in E(T)} a_e 1_{C_e} = 0$ for some coefficients a_e in $\mathbb{F}_2$. For a particular edge $e_0 \in E(T)$, we have $e_0 \in B_{e_0}$ and $e_0 \notin B_e$ for all $e \in E(T)$ such that $e \neq e_0$. So the left-hand side of $\sum_{e \in E(T)} a_e 1_{B_e} = 0$ is a_{e_0}. Thus $a_{e_0} = 0$. This proves the linear independence.

To see that $\{1_{C_e} \mid e \in E(G-T)\}$ spans the bond space of G, given a bond B of G. We claim that $1_B = \sum_{e \in E(B-T^c)} 1_{B_e}$, i.e.,

$$1_B + \sum_{e \in E(B \cap T)} 1_{B_e} = 0. \tag{2}$$

It is clear that the left-hand side of (2) is zero on all edges of $E(T-B)$. The left-hand side of (2) also cancels to zero on the edges of $E(B \cap T)$. Thus the left-hand side of (2) is the characteristic function of an edge cut U on T^c. Since $T = G - T^c$ is connected, thus $G - U$ is also connected. This implies that U must be empty. So left-hand side of (2) is identically zero. ☐

4 Flow Space and Tension Space of an Oriented Graph

Let $G = (V, E)$ be a graph. Recall that an **orientation** of G is a multi-valued function $\varepsilon : V \times E \to \{0, -1, 1\}$ such that (i) $\varepsilon(v, e) = \{\pm 1\}$ if e is a loop at v and (ii) if $e = uv$ is a non-loop at its end-vertices u, v then $\varepsilon(u, e), \varepsilon(v, e)$ are single-valued and $\varepsilon(u, e)\varepsilon(v, e) = -1$. A graph with an orientation is called an **oriented graph**. The **incidence matrix** of an oriented graph (G, ε) is a matrix $\mathbf{M} = [m_{ve}]$, where $m_{ve} = 0$ if e is a loop at v and $m_{ve} = \varepsilon(v, e)$ if e is a non-loop at v.

Let (G, ε) be an oriented graph throughout. A real-valued **flow** of (G, ε) is a function $f : E \to \mathbb{R}$ such that the flow-in equals the flow-out at every vertex v, i.e.,

$$\sum_{e \in E^-(v)} f(e) = \sum_{e \in E^+(v)} f(e), \tag{3}$$

where $E^-(v)$ and $E^+(v)$ are the sets of edges whose arrows have heads and tails at v respectively. The set of all real-valued flows of (G,ε) forms a subspace of the Euclidean $\mathbb{R}^E$, called the **flow space** of (G,ε), denoted $F(G,\varepsilon)$.

A real-valued **tension** of (G,ε) is a function $g: E \to \mathbb{R}$ such that for any directed cycle (C,ε_C) of G,

$$\sum_{e\in E(C)} [\varepsilon,\varepsilon_C](e)\, g(e) = 0.$$

The set $T(G,\varepsilon)$ of all real-valued tensions of (G,ε) forms a subspace of $\mathbb{R}^E$, called the **tension space** of (G,ε).

The inner product of $\mathbb{R}^E$ is defined by

$$\langle f,g\rangle := \sum_{e\in E} f(e)g(e).$$

Let G_i $(i=1,2)$ be subgraphs of G with orientations ε_i. The **coupling** of (G_1,ε_1) and (G_2,ε_2) is a function $[\varepsilon_1,\varepsilon_2]: E \to \{0,-1,1\}$, defined by

$$[\varepsilon_1,\varepsilon_2](e) := \begin{cases} 1 & \text{if } e\in E(G_1)\cap E(G_2) \text{ is at } v \text{ and } \varepsilon_1(v,e)=\varepsilon_2(v,e),\\ -1 & \text{if } e\in E(G_1)\cap E(G_2) \text{ is at } v \text{ and } \varepsilon_1(v,e)=\varepsilon_2(v,e),\\ 0 & \text{otherwise.}\end{cases} \tag{4}$$

If we view ε_i as the extended functions $\varepsilon_i: E \to \{0,-1,1\}$ by setting $\varepsilon_i(v,e)=0$ for edges $e\notin E(G_i)$ at their end-vertices v, then for any edge e at its end-vertex v,

$$[\varepsilon_1,\varepsilon_2](e) = \varepsilon_1(v,e)\,\varepsilon_2(v,e).$$

Let C be a cycle of G. A **direction** of C is an orientation ε_C of C such that the digraph (C,ε_C) has a head and a tail at every vertex of C; and (C,ε_C) is called a **directed cycle**. The **indicator function** I_C of C is the characteristic function of the edge set $E(C)$.

Proposition 4.1. *Let (C,ε_C) be a directed cycle of an oriented graph (G,ε). Then the coupling $[\varepsilon,\varepsilon_C]$ is a flow of (G,ε).*

Proof. For any vertex not in the cycle C, it is clear that (3) is satisfied. For a vertex $v\in V(C)$, let e and e' be edges of C at v such that $\varepsilon_C(v,e)=-1$ and $\varepsilon_C(v,e')=1$. It is routing to check

$$\sum_{x\in E^-(v)} [\varepsilon,\varepsilon_C](x) - \sum_{x\in E^+(v)} [\varepsilon,\varepsilon_C](x) = 0$$

for the cases: (1) $\varepsilon(v,e)=\varepsilon_C(v,e)$, $\varepsilon(v,e')=\varepsilon_C(v,e')$; (2) $\varepsilon(v,e)\neq\varepsilon_C(v,e)$, $\varepsilon(v,e')=\varepsilon_C(v,e')$; (3) $\varepsilon(v,e)=\varepsilon_C(v,e)$, $\varepsilon(v,e')\neq\varepsilon_C(v,e')$; and (4) $\varepsilon(v,e)\neq\varepsilon_C(v,e)$, $\varepsilon(v,e')\neq\varepsilon_C(v,e')$. □

Let $U=[X,X^c]$ be an edge cut of G. A **direction** of U is an orientation ε_U on U such that every arc of the digraph (U,ε_U) has its tail in X and head in X^c; and (U,ε_U) is called a **directed cut** of G.

Proposition 4.2. *Let (U,ε_U) be a directed cut of an oriented graph (G,ε). Then $[\varepsilon,\varepsilon_U]$ is a tension of (G,ε).*

Proof. For any edge e at its end-vertex v, we have

$$[\varepsilon,\varepsilon_C](e)\,[\varepsilon,\varepsilon_U](e) = \varepsilon(v,e)\,\varepsilon_C(v,e)\,\varepsilon(v,e)\,\varepsilon_U(v,e) = [\varepsilon_C,\varepsilon_U](e).$$

To see that $[\varepsilon,\varepsilon_U]$ is a tension, it suffices to show that

$$\sum_{e\in E(C)} [\varepsilon,\varepsilon_C](e)\,[\varepsilon,\varepsilon_U](e) = \sum_{e\in E(C)} [\varepsilon_C,\varepsilon_U](e) = \sum_{e\in E(C)\cap E(U)} [\varepsilon_C,\varepsilon_U](e)$$

is zero. Since C is a cycle and $U=[X,X^c]$ is a cut, the number of edges having agreed orientations with respect to ε_C and ε_U equals the number of edges having opposite orientations. Hence the above right-hand side is zero. □

Remark. The flows $[\varepsilon,\varepsilon_C]$ and tensions $[\varepsilon,\varepsilon_U]$ are orthogonal in the Euclidean space $\mathbb{R}^E$.

Theorem 4.3. *Let F be a forest of an oriented graph (G, ε). For each edge $e \in E(F^c)$, let C_e be the fundamental cycle of G and ε_e a direction of C_e such that orientations ε and ε_e agree on e. Then the set of flows $[\varepsilon, \varepsilon_e]$, where $e \in E(F^c)$, forms a basis for the flow space $F(G, \varepsilon)$. Moreover, for each flow $f \in F(G, \varepsilon)$,*

$$f = \sum_{e \in E(F^c)} f(e)\,[\varepsilon, \varepsilon_e].$$

Proof. We first show that $\{[\varepsilon, \varepsilon_e] \,:\, e \in E(F^c)\}$ is linearly independent. In fact, set

$$\sum_{e \in E(F^c)} a_e\,[\varepsilon, \varepsilon_e] = 0$$

for some real numbers a_e. Note that for a particular edge $e_0 \in E(F^c)$ and an arbitrary edge $e \in E(F^c)$, we have $[\varepsilon, \varepsilon_e](e_0) = 1$ if $e = e_0$ and $[\varepsilon, \varepsilon_e](e_0) = 0$ if $e \neq e_0$. Thus the above left-hand side at e_0 is a_{e_0}. Hence $a_{e_0} = 0$. We proved linear independence.

Let f be a flow of (G, ε). We claim that the function $g := f - \sum_{e \in E(F^c)} f(e)\,[\varepsilon, \varepsilon_e]$ is identically zero. It is clear that g is a flow of (G, ε), for it is a linear combinations of flows. Note that for $e_0, e \in E(F^c)$, we have $[\varepsilon, \varepsilon_e](e_0) = 1$ if $e = e_0$ and $[\varepsilon, \varepsilon_e](e_0) = 0$ if $e \neq e_0$. It follows that $g(e_0) = 0$, i.e., $g|_{E(F^c)} \equiv 0$. Since F is a forest, if F is not trivial (i.e. F contains some edge), then F has a leaf v incident with an edge e of F. For each edge $e \in E(F^c)$ at a leaf v, we must have $g(e) = 0$, for g is conservative at v. Continue this procedure; we see that $g|_{E(F)} \equiv 0$. So g is identically zero. □

Theorem 4.4. *Let F be a forest of an oriented graph (G, ε). For each edge $e \in E(F)$, let B_e be the fundamental bond of G and ε_e a direction of B_e such that orientations ε and ε_e agree on e. Then the set of flows $[\varepsilon, \varepsilon_e]$, where $e \in E(F)$, forms a basis for the tension space $T(G, \varepsilon)$. Moreover, for each tension $g \in T(G, \varepsilon)$,*

$$g = \sum_{e \in E(F)} g(e)\,[\varepsilon, \varepsilon_e].$$

Proof. We first show that $\{[\varepsilon, \varepsilon_e] \,:\, e \in E(F^c)\}$ is linearly independent. In fact, set

$$\sum_{e \in E(F)} a_e\,[\varepsilon, \varepsilon_e] = 0$$

for some real numbers a_e. Note that for a particular edge $e_0 \in E(F^c)$ and an arbitrary edge $e \in E(F^c)$, we have $[\varepsilon, \varepsilon_e](e_0) = 1$ if $e = e_0$ and $[\varepsilon, \varepsilon_e](e_0) = 0$ if $e \neq e_0$. Thus the above left-hand side at e_0 is a_{e_0}. Hence $a_{e_0} = 0$. We proved linear independence.

Let g be a tension of (G, ε). We claim that the function $h := g - \sum_{e \in E(F)} g(e)\,[\varepsilon, \varepsilon_e]$ is identically zero. It is clear that h is a tension of (G, ε), for it is a linear combinations of tensions. Note that for $e_0, e \in E(F)$, we have $[\varepsilon, \varepsilon_e](e_0) = 1$ if $e = e_0$ and $[\varepsilon, \varepsilon_e](e_0) = 0$ if $e \neq e_0$. It follows that $h(e_0) = 0$, i.e., $h|_{E(F)} \equiv 0$.

Now for each edge $x \in E(F^c)$, let ε_x be the orientation of the fundamental cycle C_x such that the orientations ε and ε_x agree on x. Since $h|_{E(F)} \equiv 0$, then by definition of tension,

$$\sum_{e \in E(C_x)} [\varepsilon, \varepsilon_x](e)\,h(e) = [\varepsilon, \varepsilon_x](x)\,h(x) = h(x)$$

is zero. Thus $h|_{E(F^c)} \equiv 0$. Hence h is identically zero. □

Corollary 4.5. *Let (G, ε) be an orientated graph. Then the flow space $F(G, \varepsilon)$ and the tension space $T(G, \varepsilon)$ are orthogonal complements in $\mathbb{R}^E$, and*

$$\dim F(G, \varepsilon) = |E(G)| - |V(G)| + c(G), \quad \dim T(G, \varepsilon) = |V(G)| - c(G).$$

5 Cayley's Formula

Theorem 5.1. *The number of labeled trees on n vertices is n^{n-2}.*

Proof. Recall that a **labeled branching** is an oriented rooted tree such that there is exactly one edge tail at each vertex, except the root. We show that the number of labeled branchings on n vertices is n^{n-1}. Then Cayley's formula follows directly because each labeled tree gives rise to n labeled branchings, one for each choice of the root vertex.

Note that each labeled branching on n vertices can be build up, one at a time to add an edge, starting with the empty graph on n labeled vertices. In order to end up with a branching, the subgraph constructed at each stage must be a branching forest (each of its component is a branching). Initially, this branching forest has n components, each consists of a isolated vertex. At each stage, we add a new edge joining a root of one branching to a vertex of another branching; the number of components decreases by one. If there are k components, the number of ways to add a new edge $e = uv$ is $n(k-1)$: each of the n vertices may be the vertex u, there are n choices for u; whereas v must be the root of a branching that does not contain the vertex u, there are $k-1$ choices for v. The total number of ways of constructing a branching on n vertices in this way is thus

$$\prod_{k=2}^{n} n(k-1) = (n-1)!\, n^{n-1}.$$

On the other hand, any individual branching on n vertices is constructed exactly $(n-1)!$ times by this procedure, once for each of the order of $n-1$ added edges. It follows that the number of labeled branchings is n^{n-1}. $\square$

Proposition 5.2. *Let G be a graph and e be a non-loop edge. Let $t(G)$ denote the number of spanning trees of a labeled graph G. Then*

$$t(G) = t(G \backslash e) + t(G/e).$$

Proof. The spanning trees of G that do not contain the edge e are exactly the spanning trees of $G\backslash e$. The spanning trees T that contain the edge e correspond to the spanning trees T/e of G/e. $\square$