08. Maximum Flow Formulations and Bipartite Matching

CPSC 535 $\sim$ Spring 2019

Kevin A. Wortman

CALIFORNIA STATE UNIVERSITY FULLERTON

October 14, 2019

Big Idea: Problem Reduction

problem A reduces to problem B = can use an algorithm for B to do all the hard work of solving problem A
= A is easier than B (or tied)

Sometimes A, B are closely related
e.g. $A =$ sorting bounded integers, $B =$ general sorting

More interesting: problems seem completely unrelated (e.g. SAT, CLIQUE; max-flow, bipartite matching)

Reduction Algorithm Pseudocode

problem A reduces to problem B = can use an algorithm for B to do all the hard work of solving problem A

```
function SOLVE-A(input-for-A)
    input-for-B = pre-process input-for-A
    solution-for-B = solve-B(input-for-B)
    solution-for-A = post-process solution-for-B
    return solution-for-A
end function
```

In spirit

- the **solve-B** part is complex and the bottleneck
- the **overhead** (pre-process and post-process parts) is simple and fast

Reducing to Max-Flow

maximum flow problem
input: a flow network G
output: a flow f of maximum value $|f|$

```
function SOLVE-A(input-for-A)
    G' = flow network based on input-for-A
    f = SOLVE-MAX-FLOW(G')
    solution-for-A = post-process f
    return solution-for-A
end function
```

(use G' because sometimes input-for-A is already a graph G)

The fastest max-flow alg. in CLRS takes $O(|V|^3)$ time; overhead usually takes linear time; so SOLVE-MAX-FLOW is usually the bottleneck.

Max-Flow Formulation

Max-Flow Formulation: details of how an algorithm for problem A

- maps an input into a flow network G'
- recovers a solution from the flow f

Also: analyze these steps to determine whether

- overhead is $O(|V|^3) \implies$ SOLVE-MAX-FLOW is the bottleneck in SOLVE-A (usually yes)
- or, overhead is $\Omega(|V|^3)$ and is the bottleneck

Usually we only discuss these parts, and don't write out the SOLVE-A pseudocode explicitly.

A Straightforward Formulation: Evacuation

Suppose we are working with safety authorities to determine how quickly CSUF could be evacuated in a natural disaster such as a wildfire.

evacuation rate problem
input: directed graph G representing a road map of Fullerton, each edge weighted with the number of autos/hour that may travel on that road
output: the maximum number of autos/hour that could travel from CSUF to a 57 or 91 freeway onramp

(Straightforward because this is clearly about flow in a directed graph.)

A Straightforward Formulation: Evacuation

For a clear formulation, need to specify

- how to convert road map into flow network G'; needs
  - to be a directed graph
  - source s and sink t
  - non-negative capacity on each edge
  - no self-loops, antiparallel edges, or disconnected vertices
- how to decode flow f into a solution for our problem (# autos/hour evacuated)
- overhead time efficiency

A Straightforward Formulation: Evacuation

- suppose road map G has none of the taboo components (self-loops etc.)
- start with $G' = G$
- define source s in G' as the Gymnasium-Campus intersection on campus
- create new sink t in G' that represents "on either freeway;" create edges from highway onramps to t, each with capacity ∞
- after finding max-flow in G', examine flow function f to compute evacuation rate as

$$\sum_{\text{onramp vertex } o} f(o, t)$$

- overhead is $O(|V| + |E|)$, not bottleneck

Generalized Max-Flow

Goal: eliminate some of the limiting constraints of the classical max-flow problem

generalized maximum flow problem
input: a flow network $G = (V, E)$, which may contain unreachable vertices, antiparallel edges, a set $S \subseteq V$ of sources, and a set $T \subseteq V$ of sinks
output: a flow f of maximum value $|f|$

Reformulating to Eliminate Unreachable Vertices

given flow network G that may contain unreachable vertices,

- use BFS or DFS to mark every vertex that is reachable from s
- use BFS/DFS again, following edges backwards, to mark every vertex that is reachable from t
- if a vertex was not marked both times, it is redundant
- G' = induced subgraph of G with all redundant vertices removed
- to convert flow f in G' to flow in G, set flow along all redundant edges to 0
- overhead is $O(|V| + |E|)$, not bottleneck

Reformulating to Eliminate Antiparallel Edges

given a flow network G that may contain antiparallel edges,

- initially $G' = G$
- identify all antiparallel edges
- when $\exists$ antiparallel edges between vertices v_1, v_2,
  - create new vertex v' in G' between v_1, v_2
  - replace edge (v_1, v_2) with edges (v_1, v') and (v', v_2)
  - set $c(v_1, v') = c(v', v_2) = c(v_1, v_2)$
- observe that flow between v_1, v_2 is identical but antiparallel edge is eliminated
- to convert flow f in G' to equiv. flow in G : for each v' introduced above, set $f(v_1, v_2) = f(v_1, v')$
- overhead is $O(|E|)$, not bottleneck

Reformulating to Accommodate Multiple Sinks or Sources

- initially $G' = G$
- create in G' a *super-source* vertex s and *super-sink* t
- for each source $s_i \in G$, create an edge (s, s_i) in G' with capacity $c(s, s_i) = \infty$
- for each sink $t_i \in G$, create an edge (t_i, t) in G' with capacity $c(t_i, t) = \infty$
- to convert flow f in G' to equiv. flow in G : delete flow info. along any of the new edges
- overhead is $O(|V|)$, not bottleneck

Formulations for Generalized Max-Flow

From now on, we have the option of formulating problems as instances of the more general max-flow problem:

generalized maximum flow problem
input: a flow network $G = (V, E)$, which may contain unreachable vertices, antiparallel edges, a set $S \subseteq V$ of sources, and a set $T \subseteq V$ of sinks
output: a flow f of maximum value $|f|$

Segue to Bipartite Matching

So far, all our reductions to max-flow have been either straightforward flow simulations, or variations on max-flow.

Now we'll see a quite-different problem that reduces to max-flow as well.

Bipartite Matching

bipartite maximum matching
input: an undirected bipartite graph $G = (V, E)$ where $V = L \cup R$ are the parts of G
output: a matching $M \subseteq E$ where the number of matched vertices is maximum

- *bipartite:* L, R are disjoint and edges only go between L, R
- *matching:* pick edges that "pair off" two vertices; goal is to maximize #paired-off
- intuitively, L is one kind of thing and R is another kind of thing

Bipartite Matching Applications

- any scenario where there are two kinds of things that can be paired
- goal is simply maximum number of pairings
- casting for a play: $L =$ set of actors; $R =$ set of roles; edge $\{l, r\}$ exists when l could play role r
- packing leftover food (one item/container): $L =$ set of food items; $R =$ available containers; edge $\{l, r\}$ exists when food l could fit in container r
- scheduling appointments: $L =$ set of clients; $R =$ set of time slots; edge $\{l, r\}$ exists when client l could meet appointment r

Formulating Bipartite Matching as Flow

- let $G = (V, E)$ be bipartite matching instance
- create $G' = (V', E')$ with $V' = V \cup \{s, t\}$ where s, t are new source/sink
- create edges
  - $(l, r)\ \forall \{l, r\} \in E$
  - $(s, l)\ \forall l \in L$
  - $(r, t)\ \forall r \in R$
- every edge (v, w) has capacity $c(v, w) = 1$
- post-processing: edge $(l, r) \in M$ iff $f(l, r) = 1$
- observe $|V'| \in O(|V|), |E'| \in O(|E|)$, overhead is $O(|V| + |E|)$
- $\implies$ if this is correct, can solve bipartite matching in $O(|V|^3)$ time

Correctness of this Formulation

Technical details:

- *integrality theorem:* if every capacity $c(u, v) \in \mathbb{Z}$ then every $f(u, v) \in \mathbb{Z}$ and $|f| \in \mathbb{Z}$
- $\exists$ matching M with cardinality $k = |M|$ iff $\exists$ some flow f with value $k = |f|$
  - key idea: pairing two vertices in the matching adds exactly one flow from $s \rightsquigarrow t$
  - there are no opportunities for flow aside from matched vertices
- $\implies$ a maximum flow in G' corresponds to a maximum matching in G