

Cliques example in SAS

This example imports an Excel file with Stanley Cup data. A graph is created by using the variables Champion and RunnerUp. An edge exists between two teams if they played each other in the Stanley Cup finals. No Canadian team has won the Stanley Cup since 1993. This code searches for cliques in the entire graph and in the subgraphs from 1994 to the present and prior to 1994. The first proc optgraph call finds all cliques in the entire data set. The second proc optgraph call finds all cliques in the first 23 observations (1994-2017). The final proc optgraph call finds all cliques prior to 1994. The collection of cliques is sent to the file specified by clique out.

```
/* Generated Code (IMPORT) */
/* Source File: stanley cup NHL clean for graph.xlsx */
/* Source Path: /gpfs/user_home/jdemaio/MATH 8020 */
/* Code generated on: 1/9/18, 1:55 PM */

%web_drop_table(WORK.IMPORT);

FILENAME REFFILE '/gpfs/user_home/jdemaio/MATH 8020/stanley cup NHL
clean for graph.xlsx';

PROC IMPORT DATAFILE=REFFILE
      DBMS=XLSX
      OUT=WORK.IMPORT;
      GETNAMES=YES;
RUN;

PROC CONTENTS DATA=WORK.IMPORT; RUN;

proc optgraph
  data_links = WORK.IMPORT;
  data_links_var
from = Champion
to = RunnerUp;
clique out = Cliques_SC_all;
* find all cliques in graph;
title 'All Cliques';
run;

proc optgraph
  data_links = WORK.IMPORT (obs=23);
  data_links_var
from = Champion
to = RunnerUp;
clique out = Cliques_SC_streak;
* find all cliques in graph since 1994;
title 'All Cliques since 1994';
run;

proc optgraph
```

```

data_links = WORK.IMPORT (firstobs=24);
data_links_var
from = Champion
to = RunnerUp;
clique out = Cliques_SC_pre_streak;
* find all cliques in graph prior to 1994;
title 'All Cliques prior to 1994';
run;

```

SAS ignores multiple edges in graphs. This will be easily seen in your log file as a warning.

Searching through the entire graph, we find cliques up to size 5. Clique #11 is the only clique of size 5.

```

23 11 New York Rangers
24 11 Boston Bruins
25 11 Detroit Red Wings
26 11 Montreal Canadiens
27 11 Toronto Maple Leafs

```

Note that proc optgraph finds maximal cliques. It will not list the five different cliques of size 4 that exist inside this clique of size 5. The only cliques of size 4 are #14 and #25.

```

34 14 Boston Bruins
35 14 Montreal Canadiens
36 14 Detroit Red Wings
37 14 Philadelphia Flyers

62 25 Detroit Red Wings
63 25 Montreal Canadiens
64 25 Chicago Black Hawks
65 25 Toronto Maple Leafs

```

Note that cliques #11 and #25 appear in the pre_streak results but #14 does not. The subclique

```

6 2 Montreal Canadiens
7 2 Boston Bruins
8 2 Philadelphia Flyers

```

does appear in the pre_streak results. The Detroit-Philadelphia edge (1997) does not exist in the pre_streak data. Only cliques of size 2 exist in streak results.

The fact that cliques of size 3 or greater exist at all is somewhat surprising. In theory, the matchup of two teams in the Stanley Cup finals is between different leagues. Hence, the resulting graph should be bipartite and have no cycles of odd length. While not common, teams can change leagues. Clearly, that happened in the NHL prior to 1994 and occurred with teams strong enough to make the Stanley Cup finals.