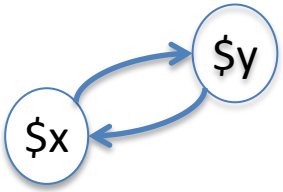
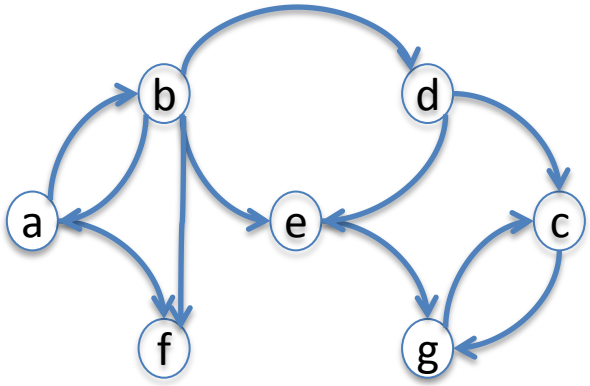
Pattern-Matching Tasks

- Find all instances of a pattern



$x = a | $x = b
$y = b | $y = a

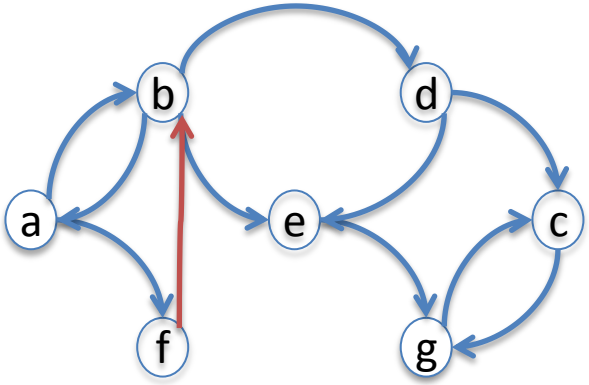
$x = c | $x = g
$y = g | $y = c



- May involve vertex or edge labels

Popular Pattern-Matching Task: Triangles

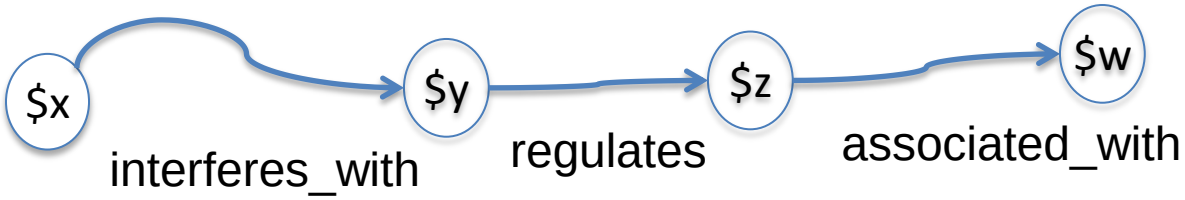
- Find Triangles a → b → c → a
- The total number of triangles is another measure of connectedness.
- Lots of algorithms; a popular challenge problem for computer scientists
- Utility in practice is not so clear (to me)
- Key idea to remember: a naïve algorithm will find the same triangle multiple times.
  - a → b → c → a
  - b → c → a → b
  - c → a → b → c
- You can extend this to k-cycles, etc.



Another Pattern Match Example

- Given a graph with edge labels
  - Drug X interferes with Drug Y
  - Drug Y regulates the expression of gene Z
  - gene Z is associated with disease w

- Find drugs that interfere with another drug involved in the treatment of a disease



Need some kind of pattern expression language

- SPARQL

SQL-like language over a fixed schema:
subject, predicate, object

```
SELECT ?x WHERE
?x interferes_with ?y .
?y regulates ?z .
?z associated_with?w .
```

subject	predicate	object
terazine	interferes_with	betamin
betamin	regulates	GT1234
GT1234	associated_with	cancer
doratin	regulates	XY6789

RDF = “Resource Description Framework”
Defines a formal data model for (subject, predicate, object) “triples.”
A triple is just a labeled edge in a graph.

Need some kind of pattern expression language

- Datalog

Assume a relation R(subject, predicate, object)

```
Ans(x) :-
    R(x, interferes_with, y),
    R(y, regulates, z),
    R(z, associated_with, w)
```

Need some kind of pattern expression language

- SQL

Assume a relation R(subject, predicate, object)

```
SELECT i.subject
FROM R i, R r, R a
WHERE i.predicate = “interferes_with”
AND r.predicate = “regulates”
AND a.predicate = “associated_with”
AND i.object = r.subject
AND r.object = a.subject
```