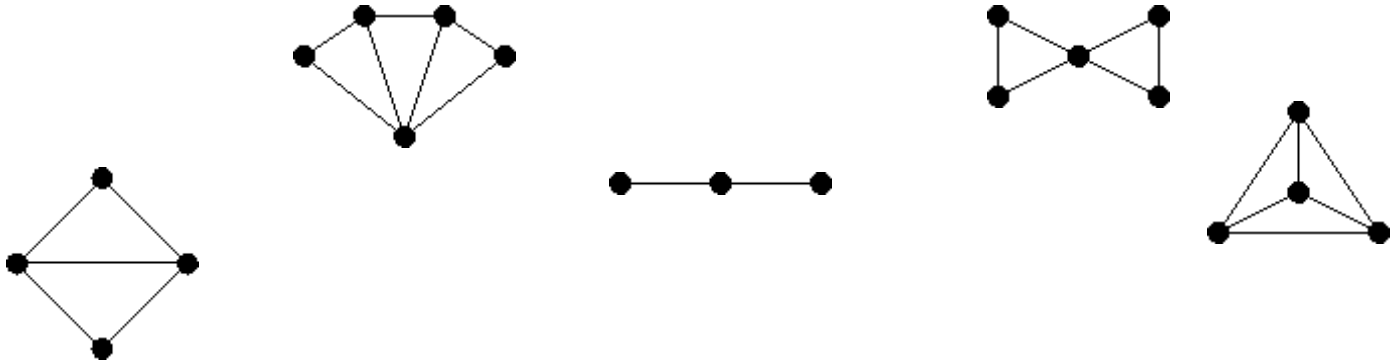
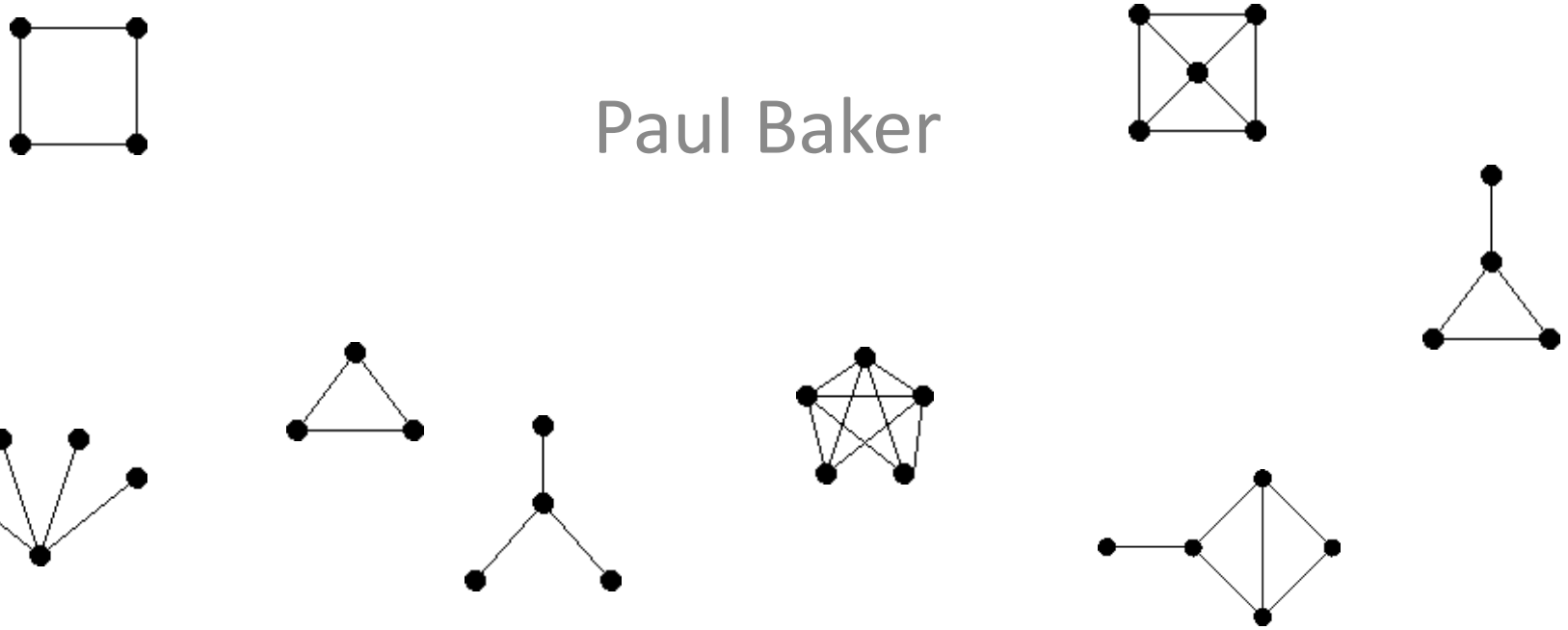
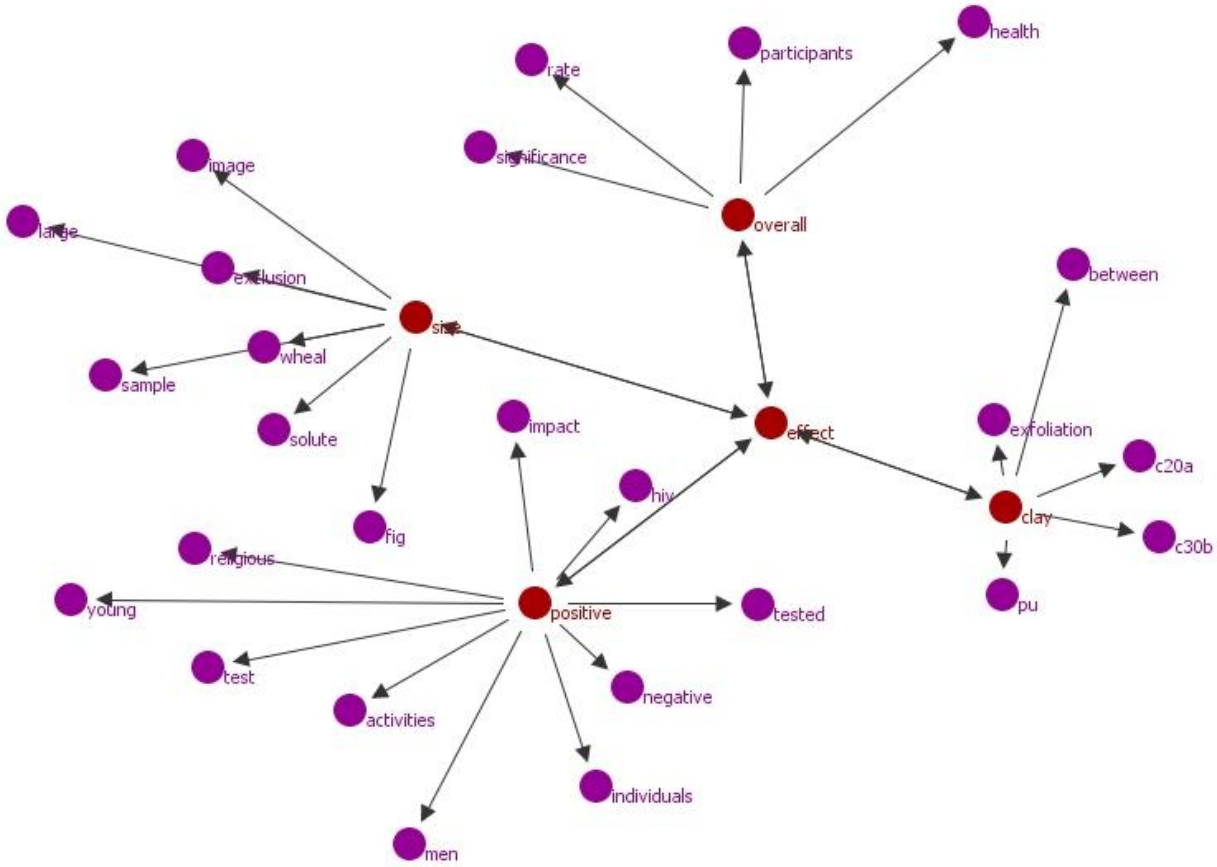


The shapes of collocation



Paul Baker

GraphColl (Brezina et al)



- 1. GraphColl compared to traditional corpus tools
- 2. Graph Theory
- 3. An experiment with different graph types

An initial research aim

- RQ: How are social actors represented in articles about Muslims in The Sun newspaper in 2010?
- Use frequency list to find candidate social actors.
- Top 20 are: *people, muslim, man, police, muslims, family, government, women, john, woman, team, wife, cameron, cops, taliban, president, choudary, dad, troops, secretary*
- This analysis focusses on *troops*.

Traditional collocational analysis

Collocate of troops	Frequency of collocate	MI score
Afghanistan	46	8.65
British	49	7.67
our	48	7.20

Min collocational frequency > 20 MI > 6 following Durrant and Doherty (2010: 145)

Results of concordancing

- *Afghanistan* + *troops*: 90% refer to troops in, killed or fighting in Afghanistan
- *British/our* + *troops* – straightforward modifier

- I AM disgusted by the vile rants and display of hatred shown towards **our** brave **troops** now serving in Afghanistan. The people involved in burning the giant poppy on Armistice Day should be deported. (The Sun, November 16, 2010)
- CHAMP TO CHUMP; Muslim who abused **our troops** is ex-British boxing title holder (The Sun, June 19, 2010)

British troops

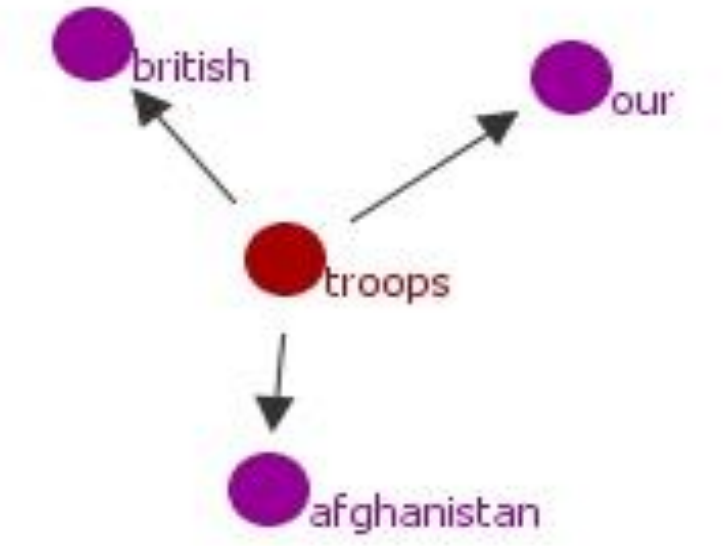
- Both belonged to a gang which gloated over terror bombings and urged the murder of **British troops** in Iraq and Afghanistan. (The Sun, January 11, 2010)
- **British troops** were also criticised by US chiefs for what they called a failure to impose security in Afghanistan. (The Sun, November 29, 2010)

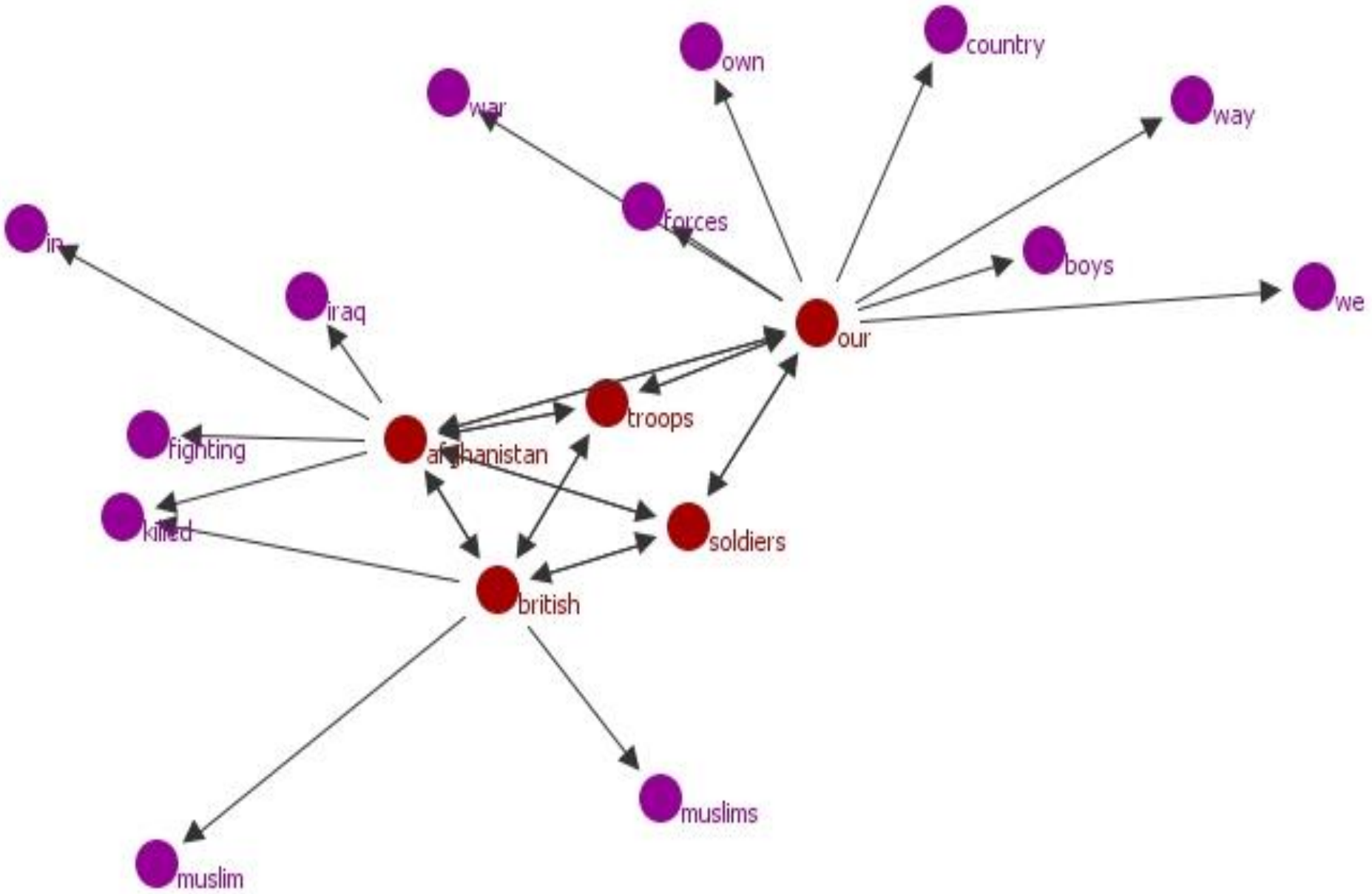
Conclusion

- *Troops* are referred to as *British* and *our*
- They are generally supported by The Sun
- They are represented as brave but under attack (unfairly) from various sources.

GraphColl analysis

GraphColl analysis




own
country
war
way
forces
in
boys
we
iraq
our
troops
fighting
afghanistan
killed
soldiers
british
muslims
muslim

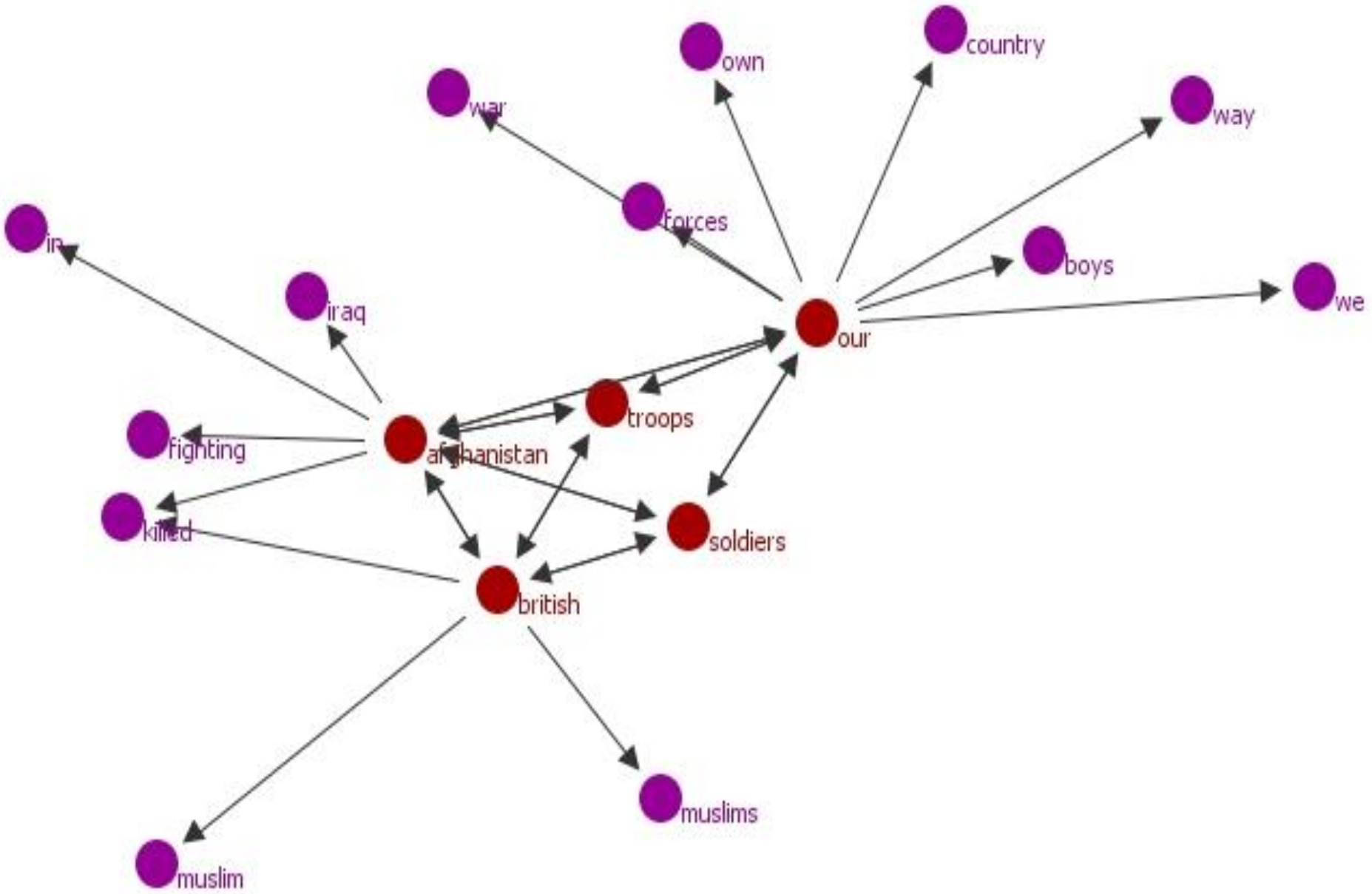
our + forces

- FIRING at civilians 21 times in four years equates to about five times a year and shows remarkable restraint from **our forces**. (The Sun, November 2, 2010)
- In Britain, on Remembrance Day when we give thanks to our war heroes, jeering fanatics hurl insults at **our forces** while police let them. (The Sun, November 16, 2010)

our + boys

- The Sun saw for itself just what **Our Boys** have been up against when we joined one of the last foot patrols by 40 Commando - in the very centre of the town which boasts a population of 20,000. (The Sun, September 21, 2010)
- **OUR Boys** are in high spirits after successfully pulling off the largest helicopter assault in British military history. (The Sun, February 15, 2010)

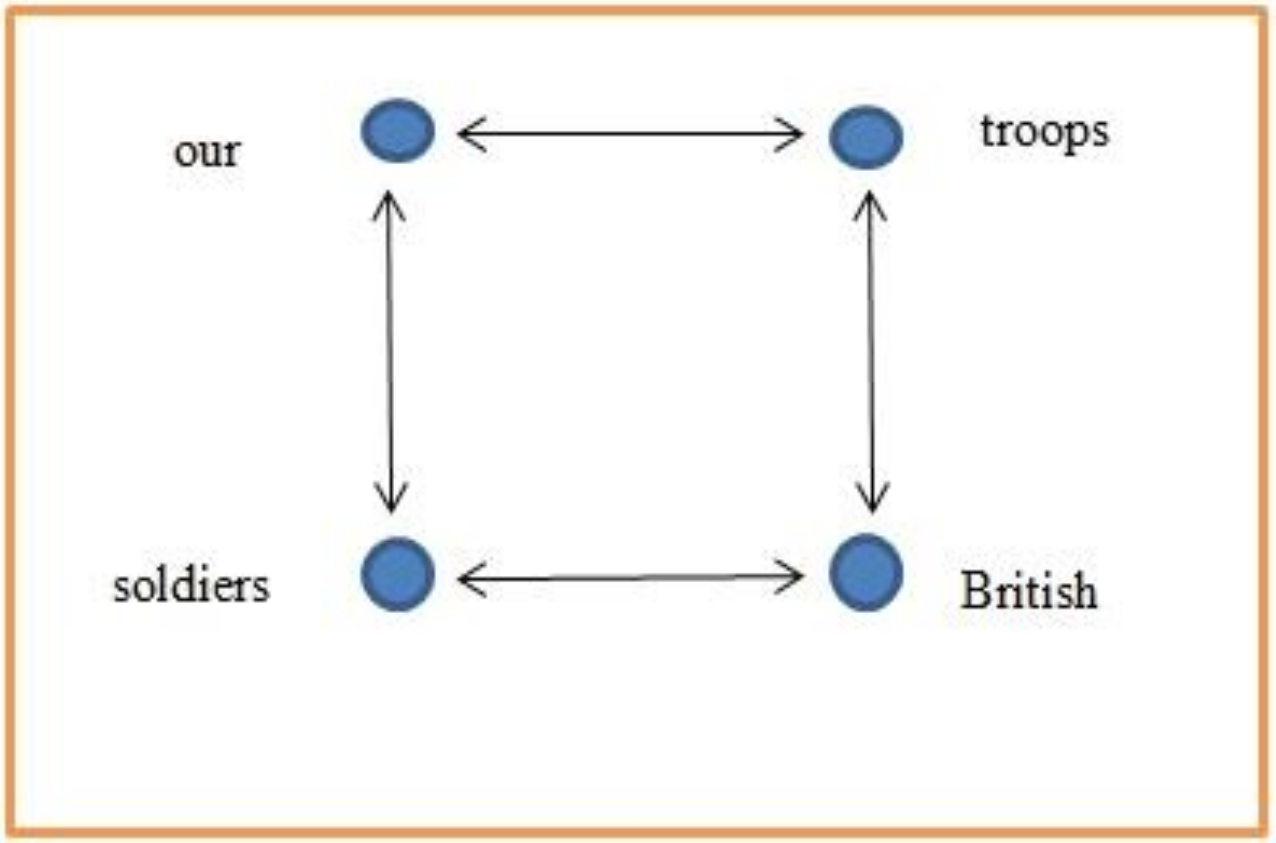
Term	Frequency
British troops	41
British soldiers	42
British boys	0
British forces	10
our troops	33
our soldiers	12
our boys	53
our forces	14


own
country
war
way
forces
in
boys
we
iraq
our
troops
fighting
afghanistan
killed
soldiers
british
muslims
muslim

Troops and soldiers have the same collocates but do not collocate

- Evil Abdul Ghani Baradar, 42 - who has the blood of 261 **British soldiers** on his hands - was tracked down by the CIA and Pakistani intelligence after FLEEING Afghanistan. (The Sun, February 17th, 2010)
- Anjem Choudary may despise this country and all it stands for but that doesn't stop him trousering an obscene amount of taxpayers' money. He actually receives £8,000 a year MORE in handouts than many **British soldiers** earn risking their lives in Afghanistan. (The Sun, Janaury 9, 2010)

our and *British* have the same collocates but do not collocate



The network helps us to interpret what "our" means

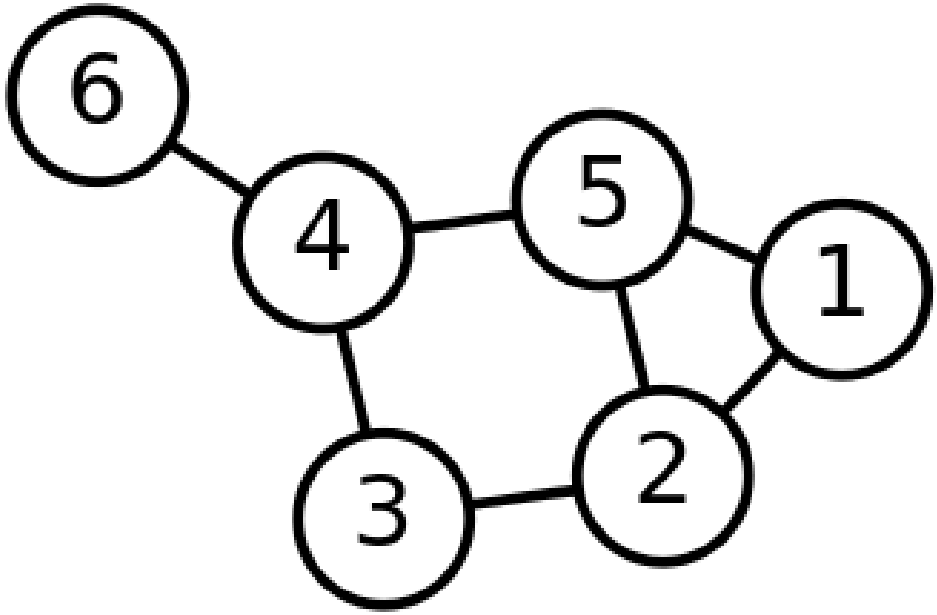
- *Our troops/soldiers/boys/forces* could be:
- The Sun's troops
- The Sun + its reader's troops
- Everyone in Britain's troops

Collocational networks

- 1: give ‘added value’ to corpus analysis by indicating relationships between multiple words which can help to suggest equivalencies, synonyms, rewordings or related terms and concepts, which (in the case of a discourse-based analysis) may have ideological significance.
- 2: They can also help to suggest relevant terms which may not have been considered for analysis in the first instance (in this case the terms *boys*).

An introduction to graph theory

- Graphs are made up of vertices (nodes or points) and arcs or lines that connect them.



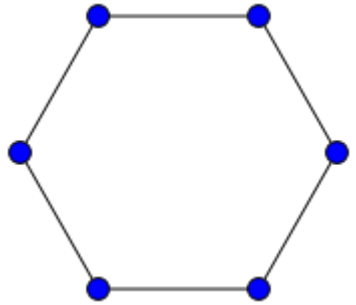
Graph Theory References

- West, D. B. *Introduction to Graph Theory, 2nd ed.* Englewood Cliffs, NJ: Prentice-Hall, p. 12, 2000.
- Harris, J. M. (2000) *Combinatorics and Graph Theory*. New York: Springer-Verlag.
- Brandstädt, A.; Le, V. B.; and Spinrad, J. P. *Graph Classes: A Survey.* Philadelphia, PA: SIAM, p. 18, 1987.
- http://mathworld.wolfram.com/
- http://www.graphclasses.org/smallgraphs.html

Some simple types of graphs

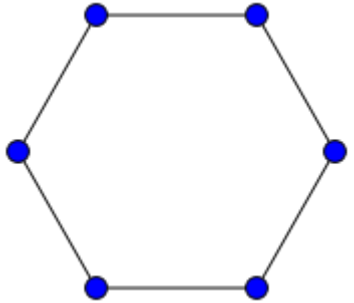
Some simple types of graphs

Cycle graph (C_6)

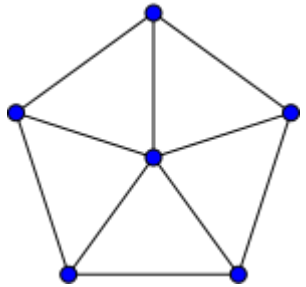


Some simple types of graphs

Cycle graph (C_6)

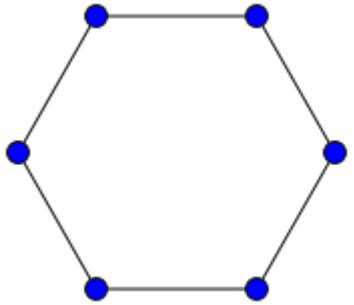


Wheel graph (W_6)

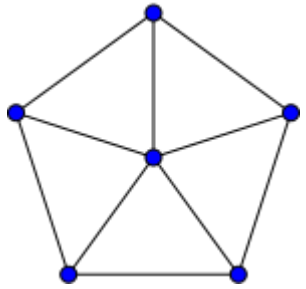


Some simple types of graphs

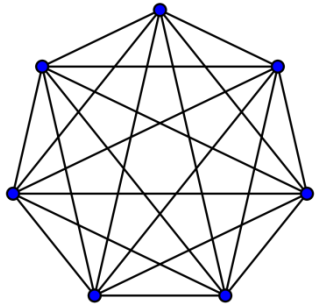
Cycle graph (C_6)



Wheel graph (W_6)

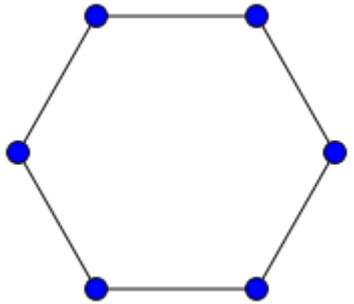


Complete graph (K_7)

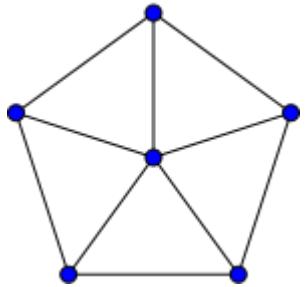


Some simple types of graphs

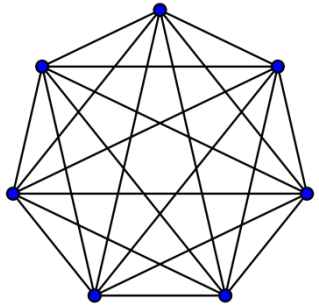
Cycle graph (C_6)



Wheel graph (W_6)

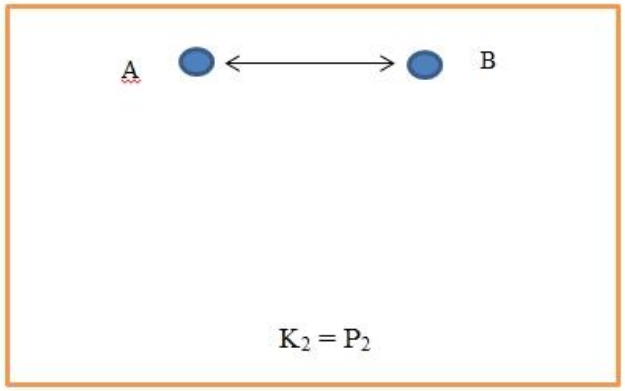


Complete graph (K_7)

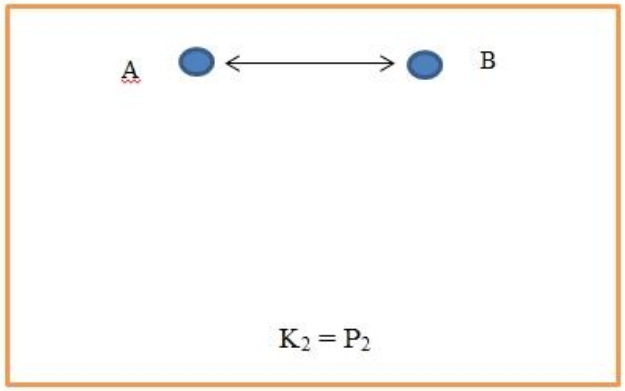


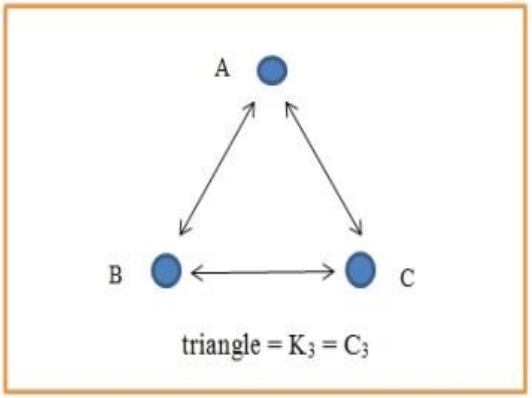
Path graph (P_6)

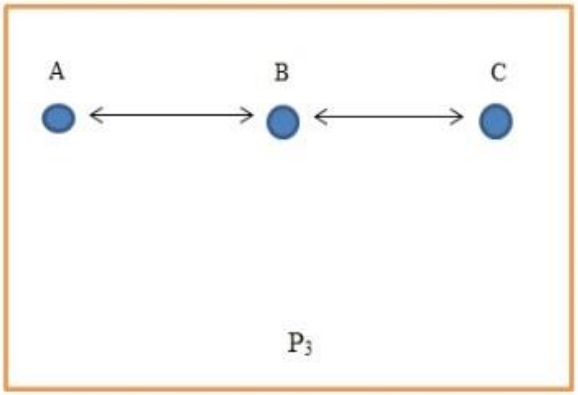
Graphs with 2, 3 or 4 nodes



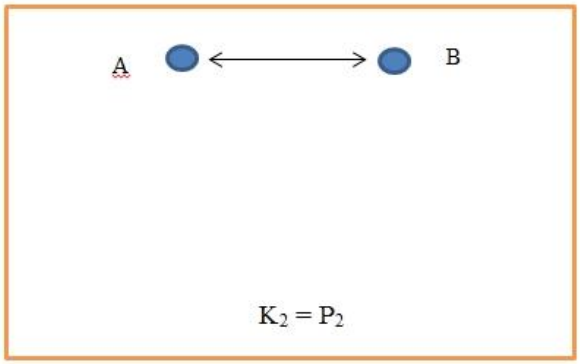
Graphs with 2, 3 or 4 nodes

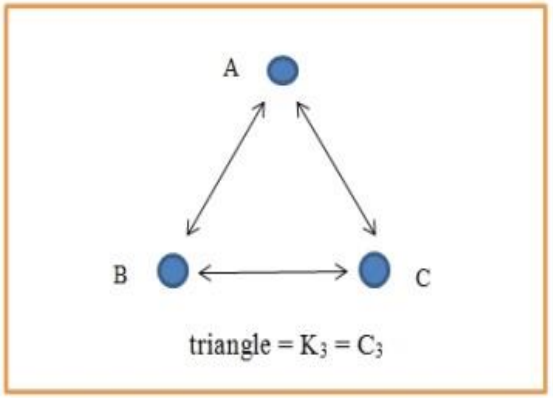


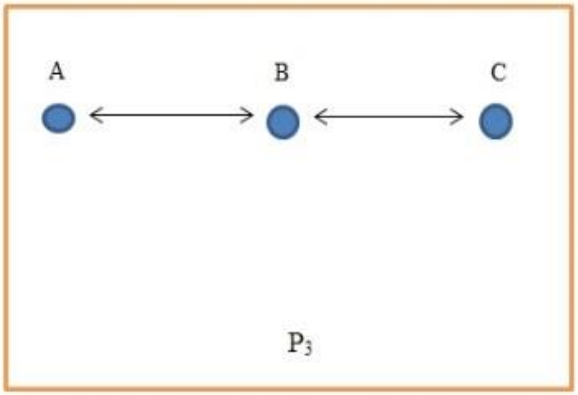


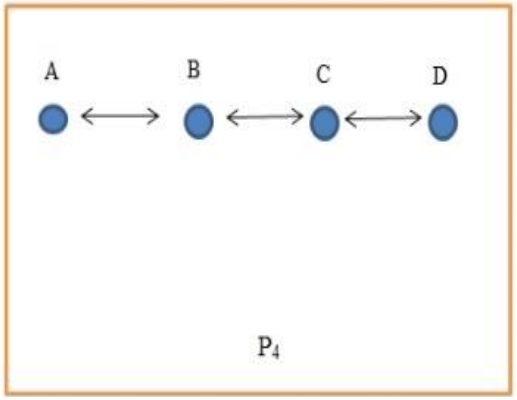


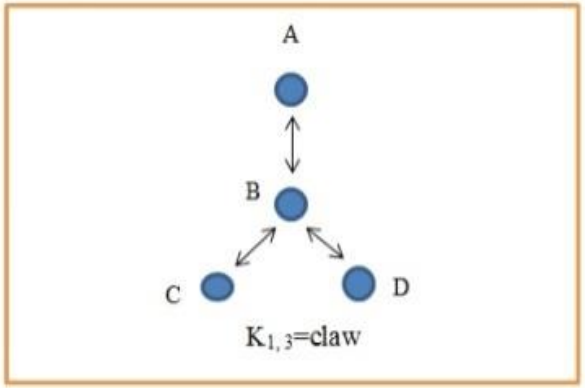
Graphs with 2, 3 or 4 nodes

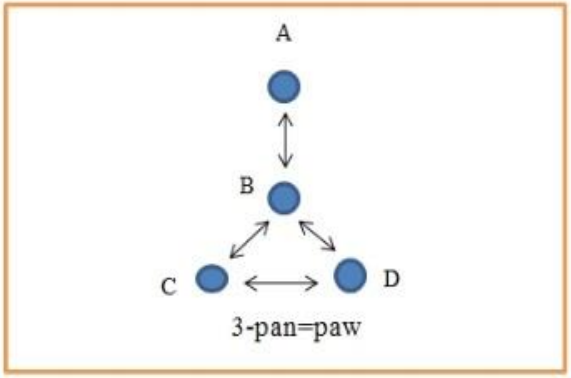


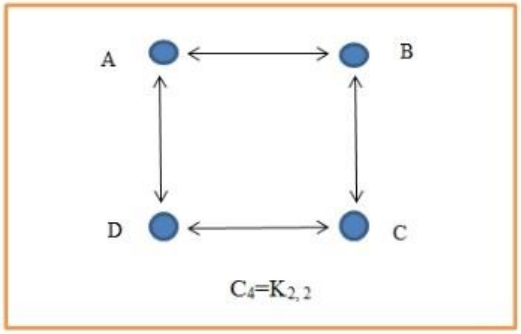


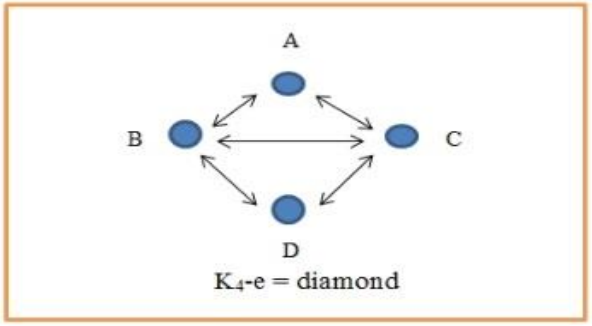


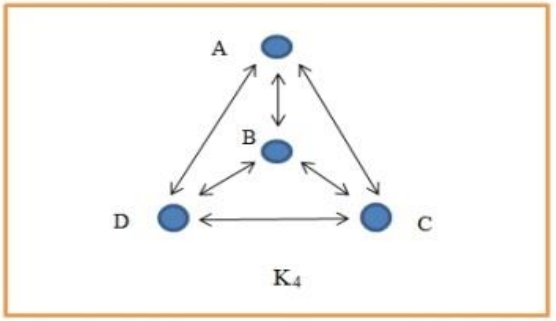




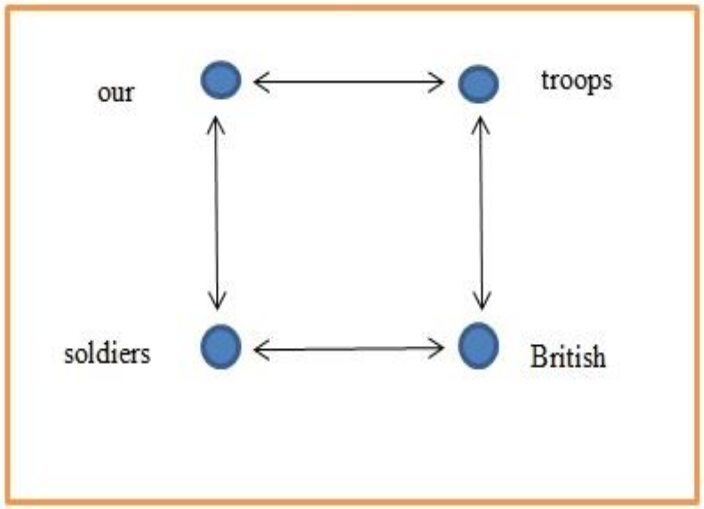


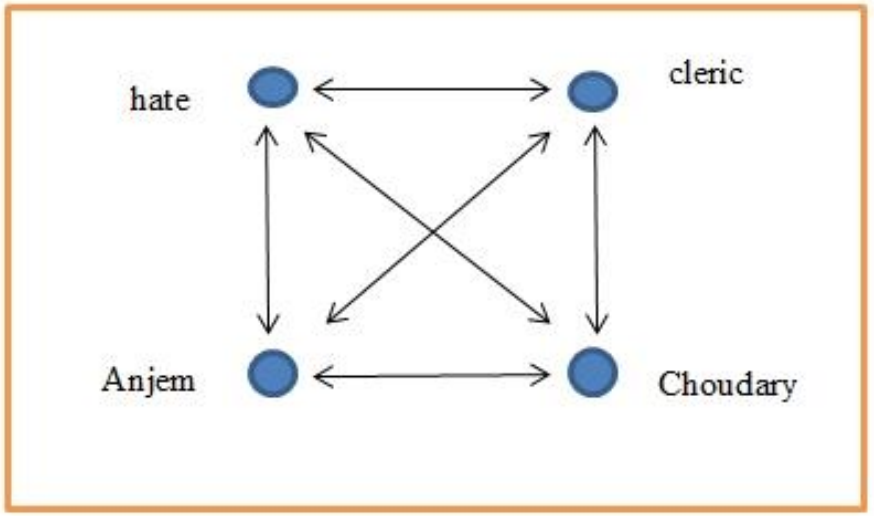






Do different graphs suggest different relationships?





- C_4 – opposite nodes suggest synonyms
- K_4 - a lexical bundle

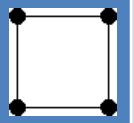
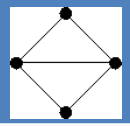
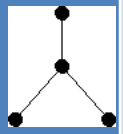
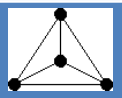
An experiment

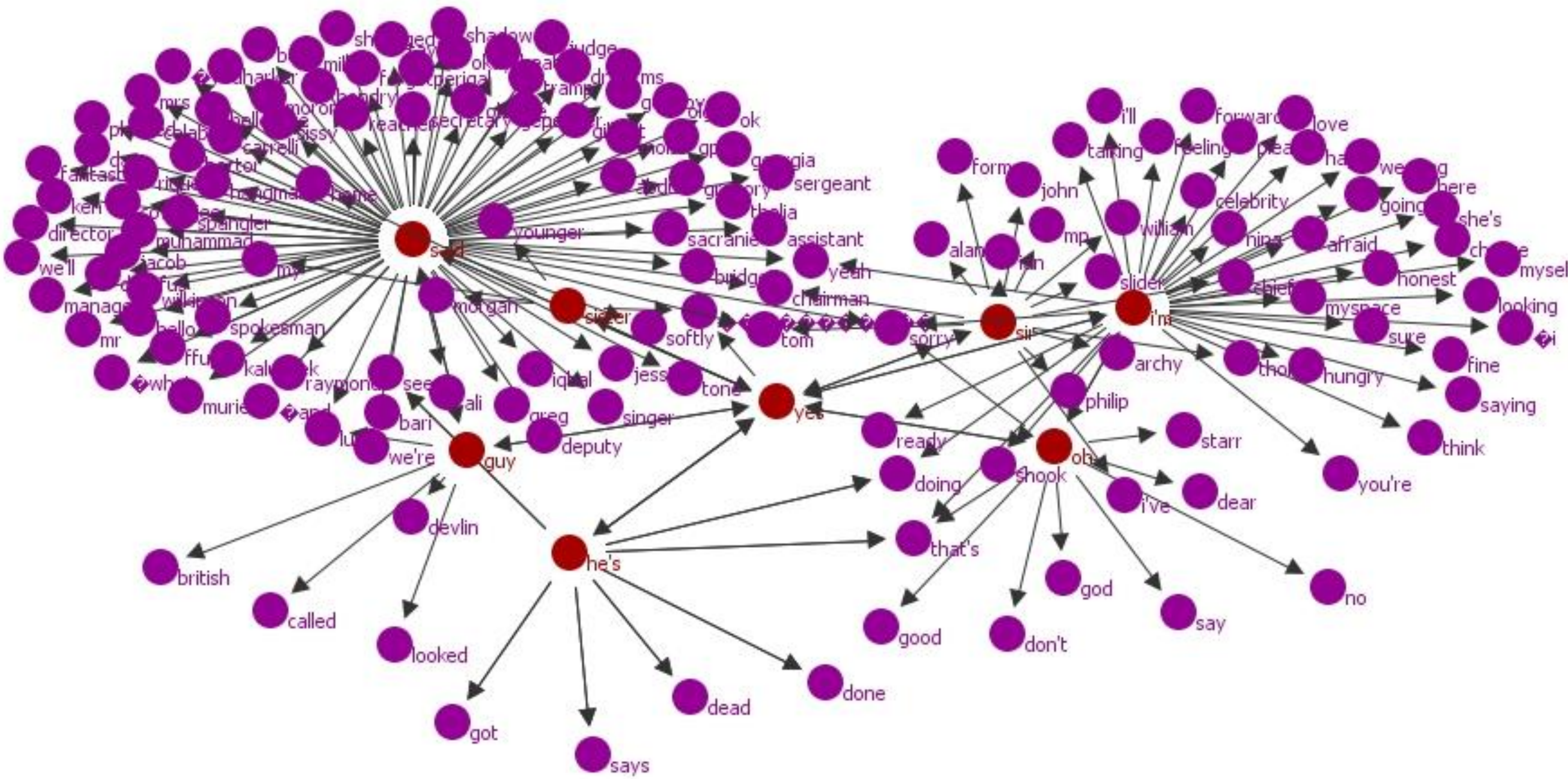
- The BE06 corpus (1 million words of written British English from 15 registers from 2006)
- I picked 40 "node" words – the 25th, 50th, 75th, 100th etc most frequent.

  you, so, other, get, day, each, et, great, help, child, full, you're, music, whole, behind, play, light, effect, yes, pay, makes, areas, account, lives, material, involved, compared, specific, costs, worked, seven, james, talking, reached, aged, shall, forces, ensure, concerned, suggest.
- I got their first and second order collocates.
- And examined the graphs to collect cases of shapes.

Cut-offs and settings

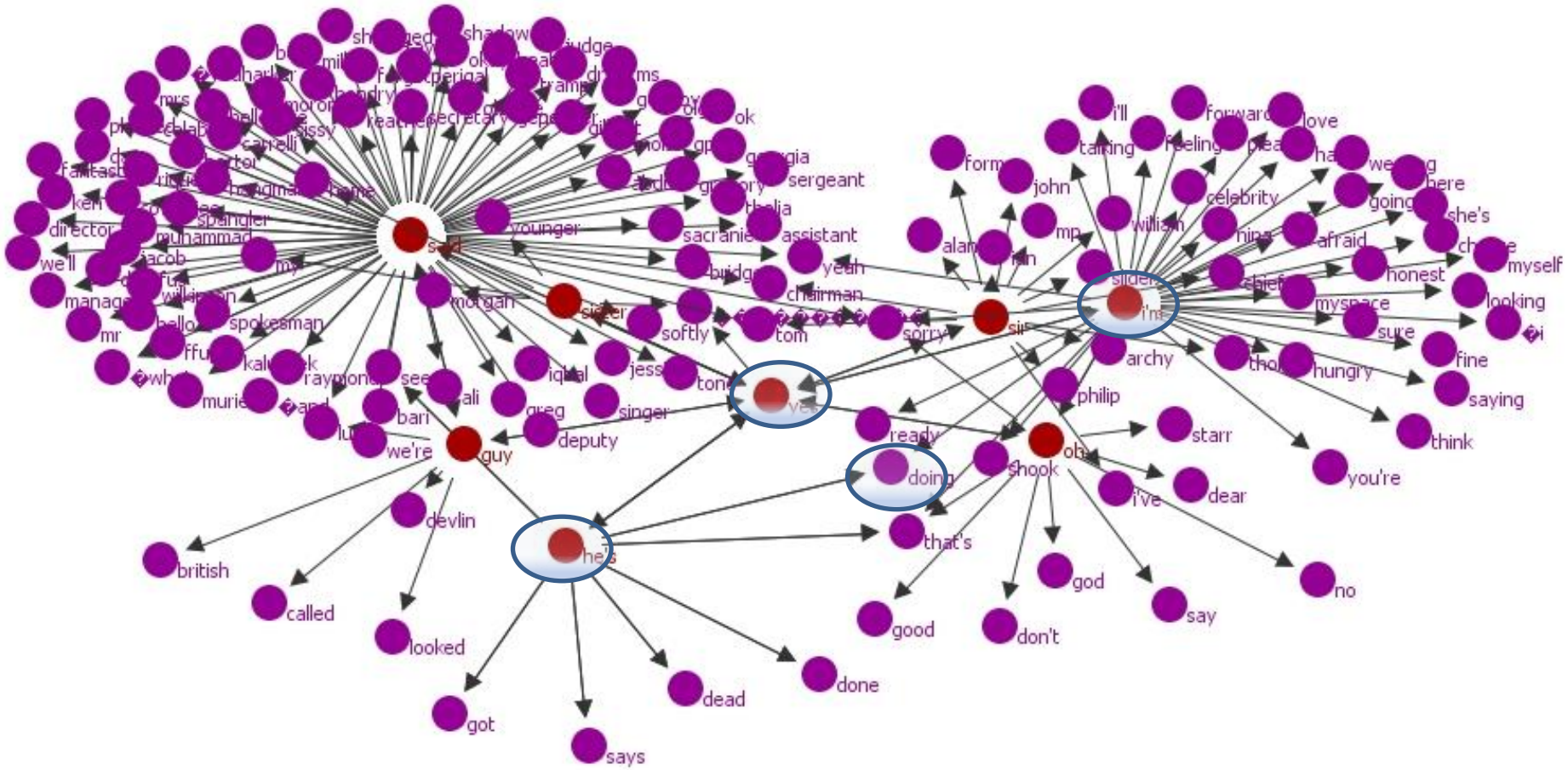
- MI > 6 again, span 5 words either side, minimum frequency = 5. If this resulted in more than 10 collocates of the node, I raised the minimum frequency until there were just 10 collocates.

Graph	#	Collection criteria
Triangle 	9	Graph must contain the node and two first order collocates (that do not collocate with any other first order collocates)
P_3 	14	Graph must either consist of a node and two of its collocates (that do not connect to anything else), or a node and any one collocate that only has one other collocate, or a node with only one first order collocate and then one of its second order collocates
C_4 	15	Graph contains node, plus two collocates which do not connect to each other, but do connect to a fourth word which is not a collocate of the node
Diamond 	15	Graph contains node and at least two of its first order collocates; everything in the graph connects, except for any two words.
Claw 	9	Graph must contain a first order collocate which collocates with the node and only two other collocates (which do not connect to each other or the node).
K_4 	3	Graph contains node and three other collocates that all collocate with one another

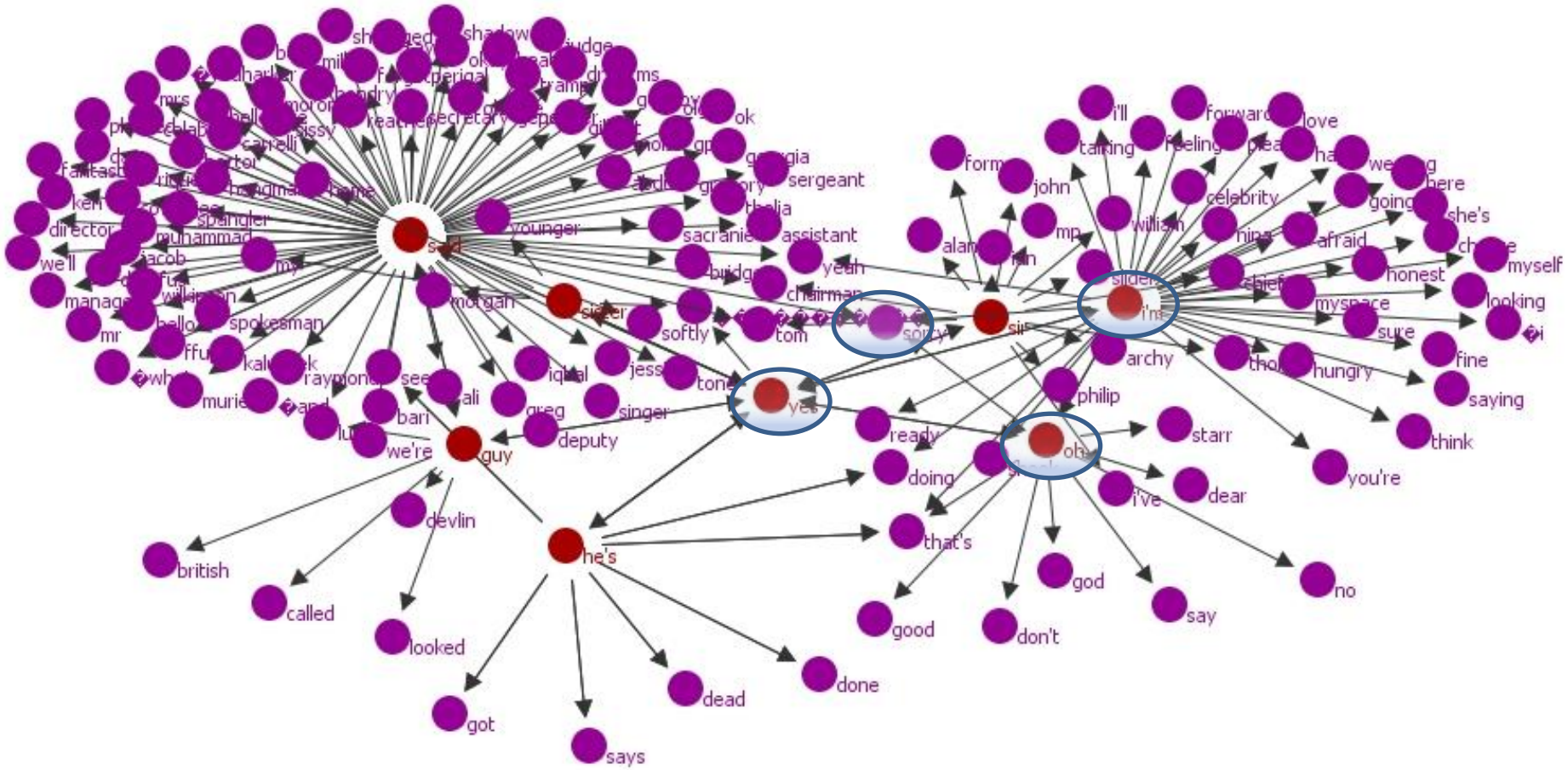


said
singer
guy
yet
he's
sir
i'm
oh
younger
sacranie
assistant
sergeant
georgia
yeah
chairman
softly
tom
sorry
morgan
iqbal
singer
deputy
greg
ali
bari
we're
devlin
british
called
looked
got
says
dead
done
that's
doing
ready
good
don't
god
say
no
dear
starr
i've
shook
philip
archy
slider
john
mp
alan
william
celebrity
i'll
talking
feeling
forward
love
here
she's
going
afraid
honest
myself
myspace
looking
sure
fine
hungry
saying
think
you're
spokesman
muhammad
spangler
director
manager
mr
we'll
jacob
mrs
carrelli
raymond
secretary
shadow
judge

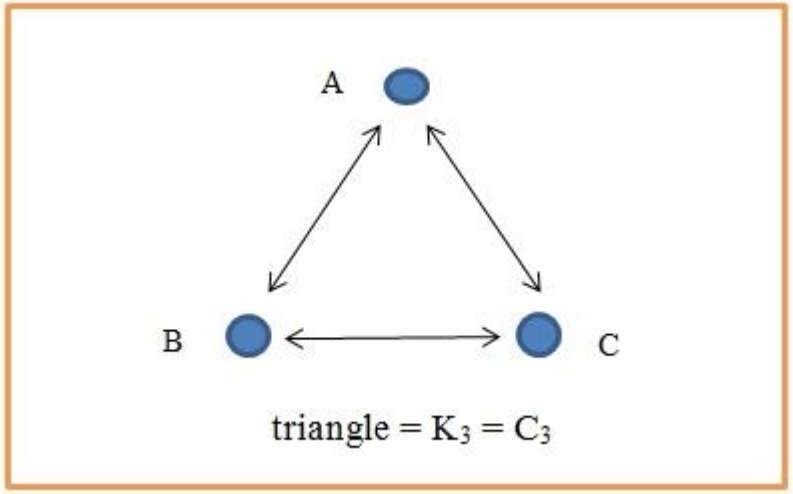
C4: yes, he’s, I’m and doing



diamond (sorry, I’m, yes, oh).

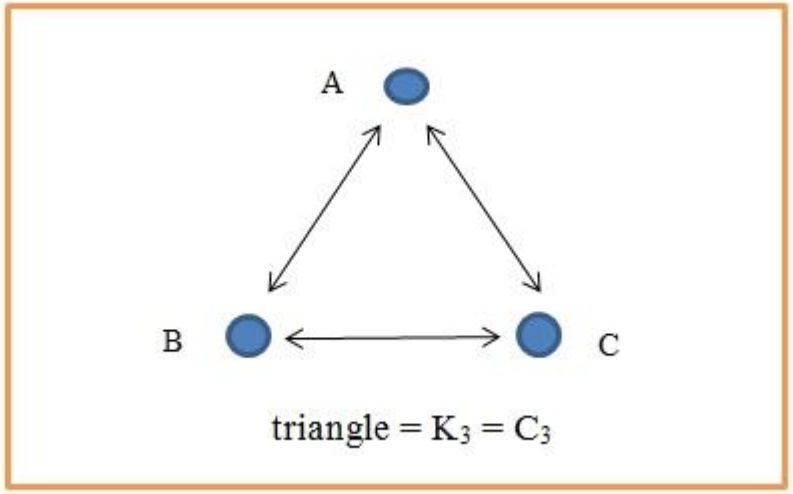


Triangles



	A	B	C
1	child	parental	leave
2	you're	I'm	going
3	music	laptop	live
4	women	men	compared
5	costs	total	per
6	james	hellebore	butcher
7	Britain's	armed	forces
8	studies	results	suggest
9	you	don't	know

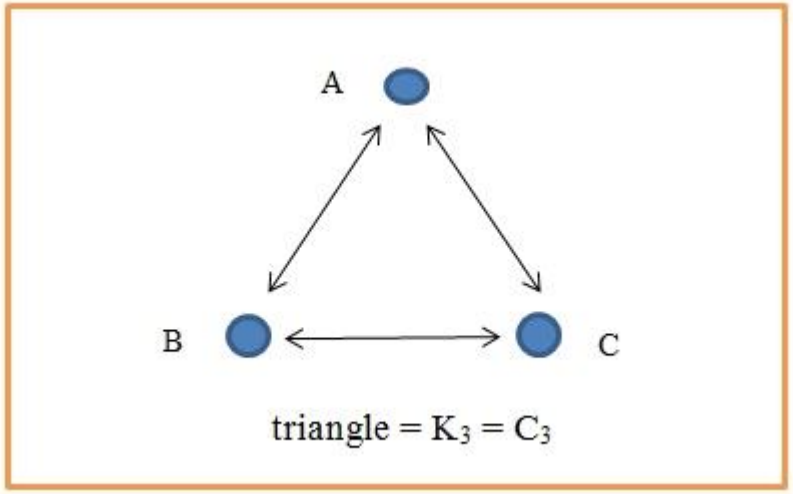
Triangles



4 out of 9 have 2+ words are from related categories (POS/semantic)

	A	B	C
1	child	parental	leave
2	you're	I'm	going
3	music	laptop	live
4	women	men	compared
5	costs	total	per
6	james	hellebore	butcher
7	Britain's	armed	forces
8	studies	results	suggest
9	you	don't	know

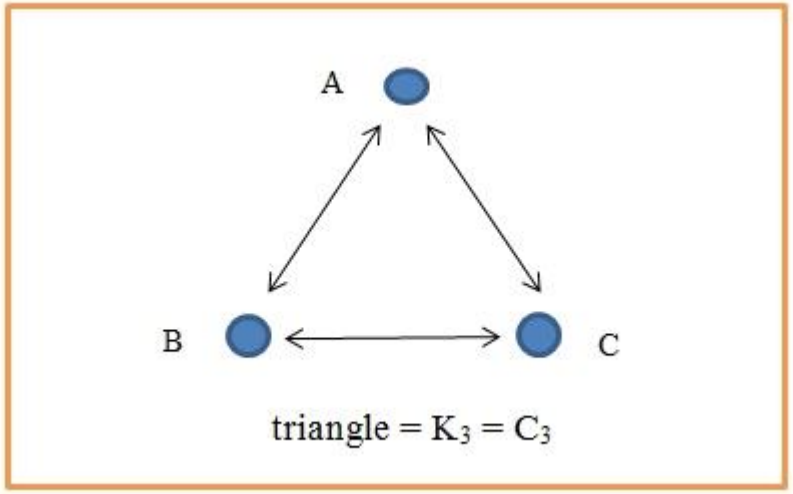
Triangles



6 out of 9 contain at least 1 2-word bundle

	A	B	C
1	child	parental	leave
2	you're	I'm	going
3	music	laptop	live
4	women	men	compared
5	costs	total	per
6	james	hellebore	butcher
7	Britain's	armed	forces
8	studies	results	suggest
9	you	don't	know

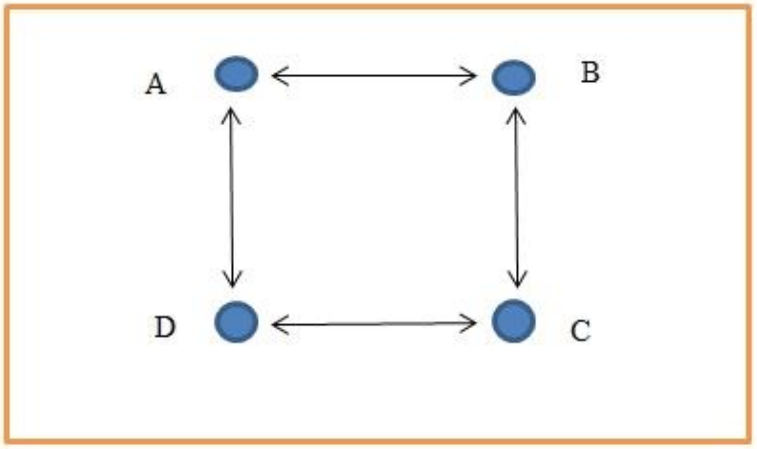
Triangles



total *eviction* costs
total *prosecution* costs
total *ward* costs

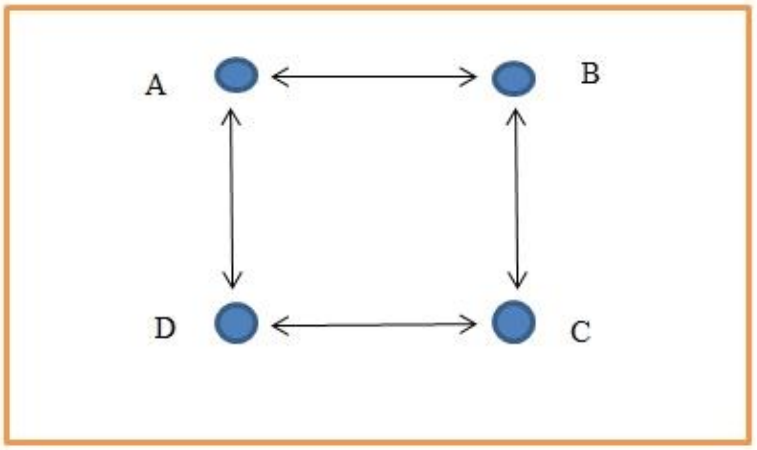
	A	B	C
1	child	parental	leave
2	you're	I'm	going
3	music	laptop	live
4	women	men	compared
5	costs	total	per
6	james	hellebore	butcher
7	Britain's	armed	forces
8	studies	results	suggest
9	you	don't	know

C_4 graphs



	A	B	C	D
1	risk	lives	people's	children
2	ethnic	specific	religious	groups
3	day	christmas	night	cold
4	can’t	help	couldn’t	tell
5	behind	turned	towards	door
6	I'm	doing	he’s	yes
7	think	makes	feel	don’t
8	taken	account	taking	steps
9	those	compared	group	pain
10	days	few	months	seven
11	I	can't	you	know
12	so	far	too	much
13	variables	other	categories	between
14	mrs	james	hellebore	said
15	children	aged	years	thousands

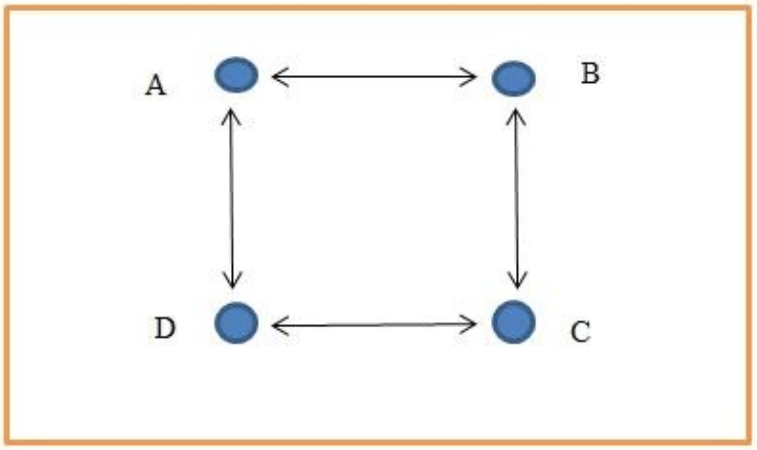
C_4 graphs



12 out of 15 graphs show relationships between non-collocating words in A-C

	A	B	C	D
1	risk	lives	people's	children
2	ethnic	specific	religious	groups
3	day	christmas	night	cold
4	can’t	help	couldn’t	tell
5	behind	turned	towards	door
6	I'm	doing	he’s	yes
7	think	makes	feel	don’t
8	taken	account	taking	steps
9	those	compared	group	pain
10	days	few	months	seven
11	I	can't	you	know
12	so	far	too	much
13	variables	other	categories	between
14	mrs	james	hellebore	said
15	children	aged	years	thousands

C_4 graphs



The 3 atypical cases contain lexical bundles:

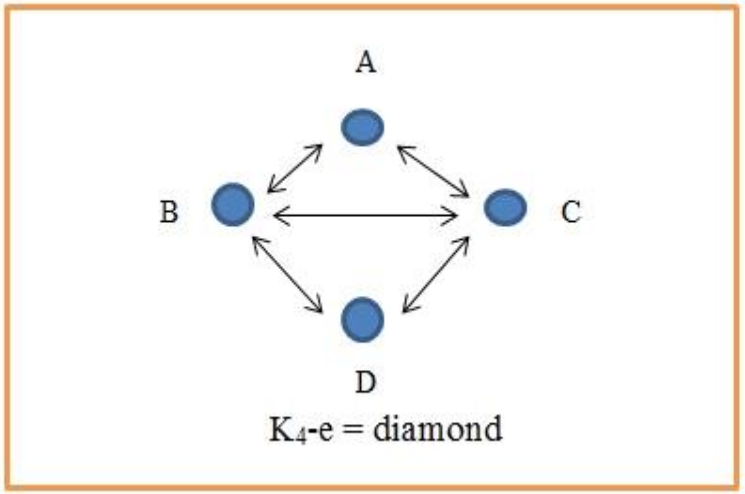
people's lives

pain group

children aged

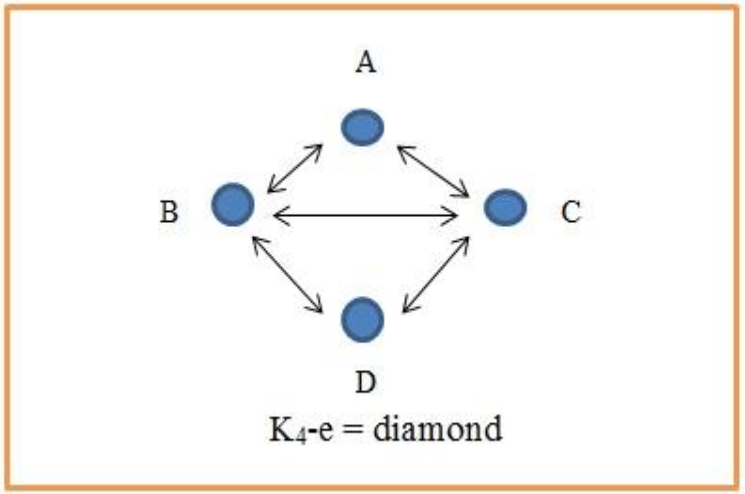
	A	B	C	D
1	risk	lives	people's	children
2	ethnic	specific	religious	groups
3	day	christmas	night	cold
4	can't	help	couldn't	tell
5	behind	turned	towards	door
6	I'm	doing	he's	yes
7	think	makes	feel	don't
8	taken	account	taking	steps
9	those	compared	group	pain
10	days	few	months	seven
11	I	can't	you	know
12	so	far	too	much
13	variables	other	categories	between
14	mrs	james	hellebore	said
15	children	aged	years	thousands

Diamonds



	A	B	D	D
1	harmonius	play	free	disharmonius
2	pay	men	sex	paid
3	bills	shall	private	bill
4	child	health	development	education
5	2001	et	al	2002
6	advice	help	information	further
7	shut	door	behind	closed
8	eyes	green	bright	light
9	bias	material	along	straining
10	compared	men	women	sex
11	total	costs	per	pounds
12	years	aged	25	per
13	muscle	get	meal	ripped
14	do	you	want	don't
15	sorry	i'm	oh	yes

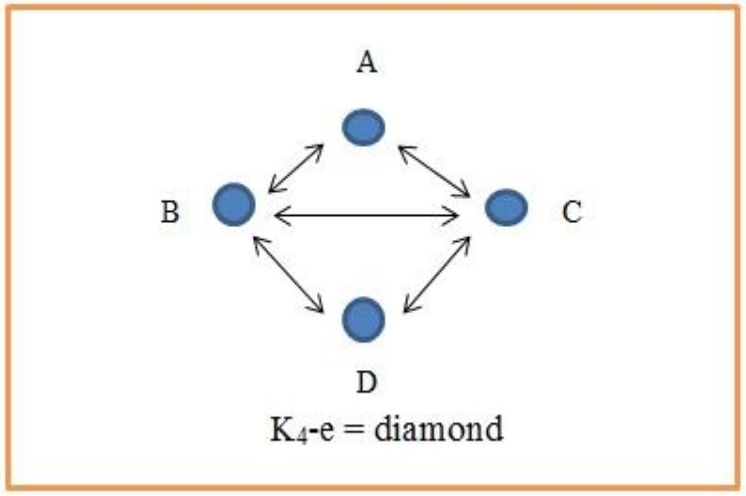
Diamonds



7 out of 15 cases indicate relationships between non-collocating A-D

	A	B	D	D
1	harmonius	play	free	disharmonius
2	pay	men	sex	paid
3	bills	shall	private	bill
4	child	health	development	education
5	2001	et	al	2002
6	advice	help	information	further
7	shut	door	behind	closed
8	eyes	green	bright	light
9	bias	material	along	straining
10	compared	men	women	sex
11	total	costs	per	pounds
12	years	aged	25	per
13	muscle	get	meal	ripped
14	do	you	want	don't
15	sorry	i'm	oh	yes

Diamonds



6 show lexical bundles

child health/development

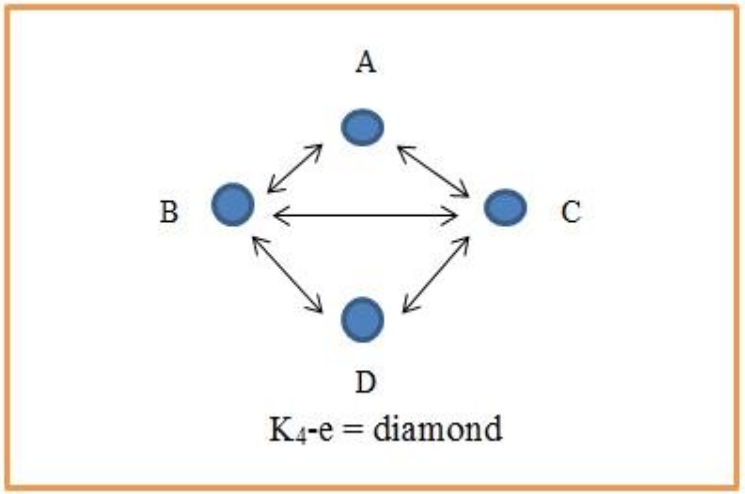
further information

help and information

information and advice

	A	B	D	D
1	harmonius	play	free	disharmonius
2	pay	men	sex	paid
3	bills	shall	private	bill
4	child	health	development	education
5	2001	et	al	2002
6	advice	help	information	further
7	shut	door	behind	closed
8	eyes	green	bright	light
9	bias	material	along	straining
10	compared	men	women	sex
11	total	costs	per	pounds
12	years	aged	25	per
13	muscle	get	meal	ripped
14	do	you	want	don't
15	sorry	i'm	oh	yes

Diamonds



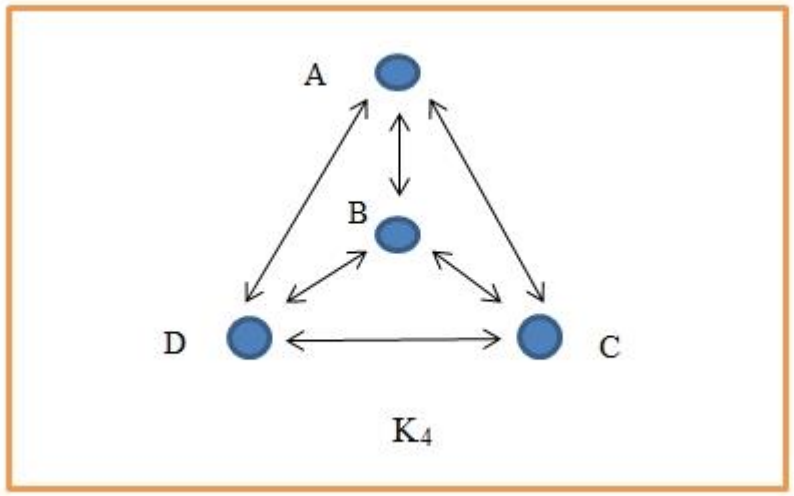
2 occur in frames:

men who [pay/had paid] for sex

get [ripped/muscle] meal plan

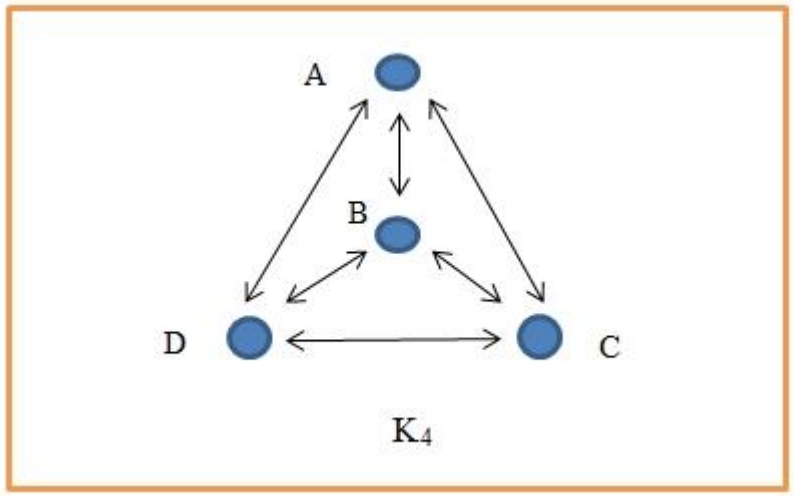
	A	B	D	D
1	harmonius	play	free	disharmonius
2	pay	men	sex	paid
3	bills	shall	private	bill
4	child	health	development	education
5	2001	et	al	2002
6	advice	help	information	further
7	shut	door	behind	closed
8	eyes	green	bright	light
9	bias	material	along	straining
10	compared	men	women	sex
11	total	costs	per	pounds
12	years	aged	25	per
13	muscle	get	meal	ripped
14	do	you	want	don't
15	sorry	i'm	oh	yes

K_4 graphs



	A	B	C	D
1	committee	shall	private	bill
2	aged	over	50	years
3	get	ripped	workout	meal

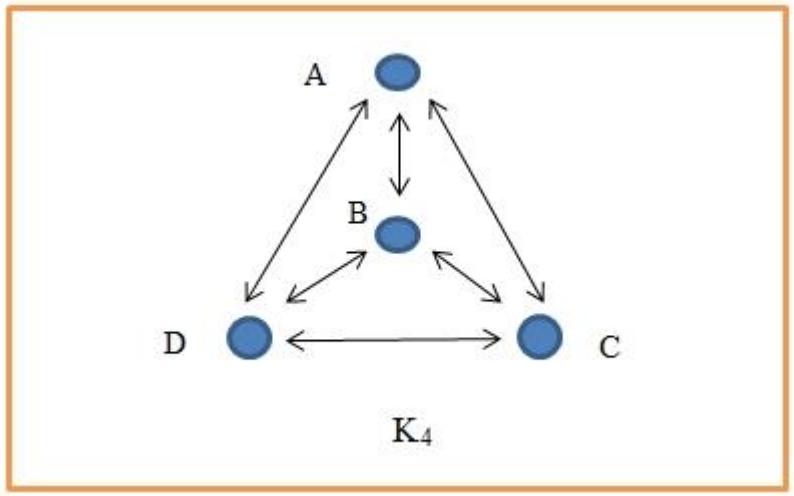
K_4 graphs



	A	B	C	D
1	committee	shall	private	bill
2	aged	over	50	years
3	get	ripped	workout	meal

*the **committee** of selection **shall** (7)*
private bill shall (5)
committee** on an opposed **private bill (6)

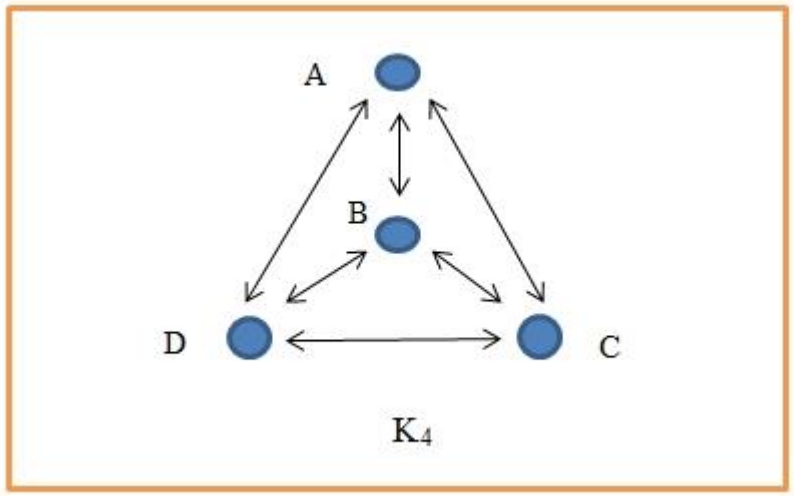
K_4 graphs



	A	B	C	D
1	committee	shall	private	bill
2	aged	over	50	years
3	get	ripped	workout	meal

50 years *(7)*
over *the* ***years*** *(19)*
aged *[under/****over****]* ***50*** *(7)*
over *the [next/past]* ***years*** *(25)*
aged *[n]* ***years*** *(18)*

K_4 graphs



	A	B	C	D
1	committee	shall	private	bill
2	aged	over	50	years
3	get	ripped	workout	meal

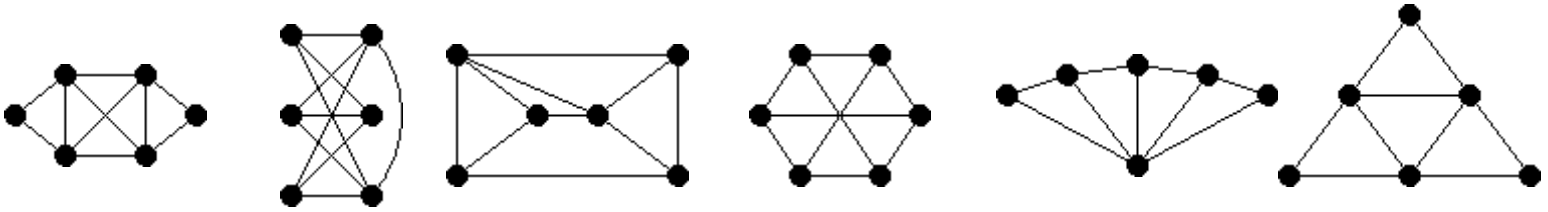
get muscle *(12)*
get muscle meal *plan (6)*
get ripped *(20)*
get ripped meal *plan (11)*
workout get (8)
ripped workout *(5)*

Conclusions

- GraphColl adds a new dimension to collocational analysis – it is a “game changer”
- The concepts and terminology around graph theory can be adopted by corpus linguists
- Some graphs are likely to be indicative of relationships or equivalencies between words, others indicate lexical bundles or frames
- Collocational networks are complicated - analysts are advised to develop skills in identifying particular graphs and recognising what they usually mean.

Further work

- Experiment with different collocational settings
- Analysis could be expanded to consider graphs with 5+ nodes



- More work to be done on how graphs link together