Biconnected components [and Articulation Points and Bridges] in MapReduce, using Graph (Navigational) Sketches

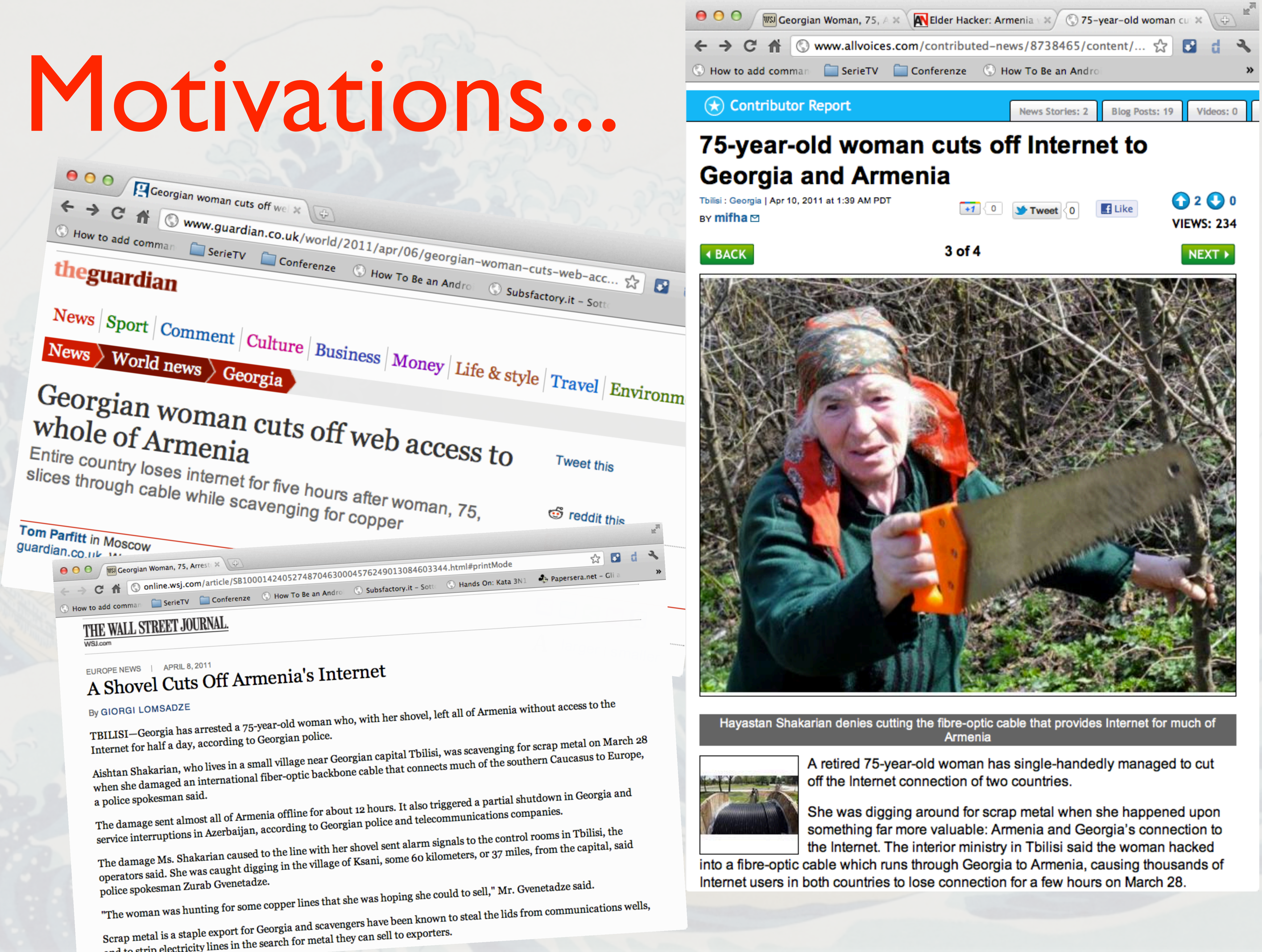
Luigi Laura,
joint work with Giorgio Ausiello,
Donatella Firmani and Emanuele Paracone

Warm up...

Let us start from the computation of (bi)connected components, articulation points and bridges in the (Semi-)Streaming Model...

...let us say we have a network, where edges are added one by one (i.e. they arrive in the stream) and we want to compute the above properties...

Motivations...



theguardian

News | Sport | Comment | Culture | Business | Money | Life & style | Travel | Environm

News › World news › Georgia

Georgian woman cuts off web access to whole of Armenia

Entire country loses internet for five hours after woman, 75, slices through cable while scavenging for copper

Tweet this

reddit this

Tom Parfitt in Moscow

THE WALL STREET JOURNAL.

EUROPE NEWS | APRIL 8, 2011

A Shovel Cuts Off Armenia's Internet

By GIORGI LOMSADZE

TBILISI—Georgia has arrested a 75-year-old woman who, with her shovel, left all of Armenia without access to the Internet for half a day, according to Georgian police.

Aishtan Shakarian, who lives in a small village near Georgian capital Tbilisi, was scavenging for scrap metal on March 28 when she damaged an international fiber-optic backbone cable that connects much of the southern Caucasus to Europe, a police spokesman said.

The damage sent almost all of Armenia offline for about 12 hours. It also triggered a partial shutdown in Georgia and service interruptions in Azerbaijan, according to Georgian police and telecommunications companies.

The damage Ms. Shakarian caused to the line with her shovel sent alarm signals to the control rooms in Tbilisi, the operators said. She was caught digging in the village of Ksani, some 60 kilometers, or 37 miles, from the capital, said police spokesman Zurab Gvenetadze.

"The woman was hunting for some copper lines that she was hoping she could to sell," Mr. Gvenetadze said.

Scrap metal is a staple export for Georgia and scavengers have been known to steal the lids from communications wells, and to strip electricity lines in the search for metal they can sell to exporters.

Contributor Report

75-year-old woman cuts off Internet to Georgia and Armenia

Tbilisi : Georgia | Apr 10, 2011 at 1:39 AM PDT

BY mifha

VIEWS: 234

BACK | 3 of 4 | NEXT

Hayastan Shakarian denies cutting the fibre-optic cable that provides Internet for much of Armenia

A retired 75-year-old woman has single-handedly managed to cut off the Internet connection of two countries.

She was digging around for scrap metal when she happened upon something far more valuable: Armenia and Georgia's connection to the Internet. The interior ministry in Tbilisi said the woman hacked into a fibre-optic cable which runs through Georgia to Armenia, causing thousands of Internet users in both countries to lose connection for a few hours on March 28.

M

THE WALL STREET JOURNAL.
WSJ.com

EUROPE NEWS | APRIL 8, 2011

A Shovel Cuts Off Armenia's Internet

By GIORGI LOMSADZE

TBILISI—Georgia has arrested a 75-year-old woman who, with her shovel, left all of Armenia without access to the Internet for half a day, according to Georgian police.

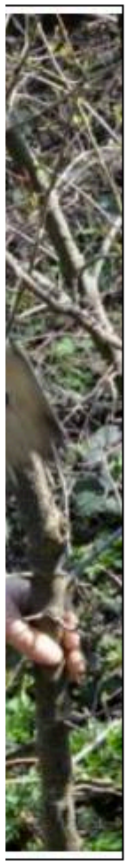
Aishtan Shakarian, who lives in a small village near Georgian capital Tbilisi, was scavenging for scrap metal on March 28 when she damaged an international fiber-optic backbone cable that connects much of the southern Caucasus to Europe, a police spokesman said.

The damage sent almost all of Armenia offline for about 12 hours. It also triggered a partial shutdown in Georgia and service interruptions in Azerbaijan, according to Georgian police and telecommunications companies.

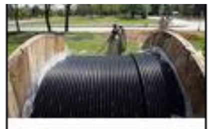
The damage Ms. Shakarian caused to the line with her shovel sent alarm signals to the control rooms in Tbilisi, the operators said. She was caught digging in the village of Ksani, some 60 kilometers, or 37 miles, from the capital, said police spokesman Zurab Gvenetadze.

"The woman was hunting for some copper lines that she was hoping she could to sell," Mr. Gvenetadze said.

Scrap metal is a staple export for Georgia and scavengers have been known to steal the lids from communications wells, and to strip electricity lines in the search for metal they can sell to exporters.



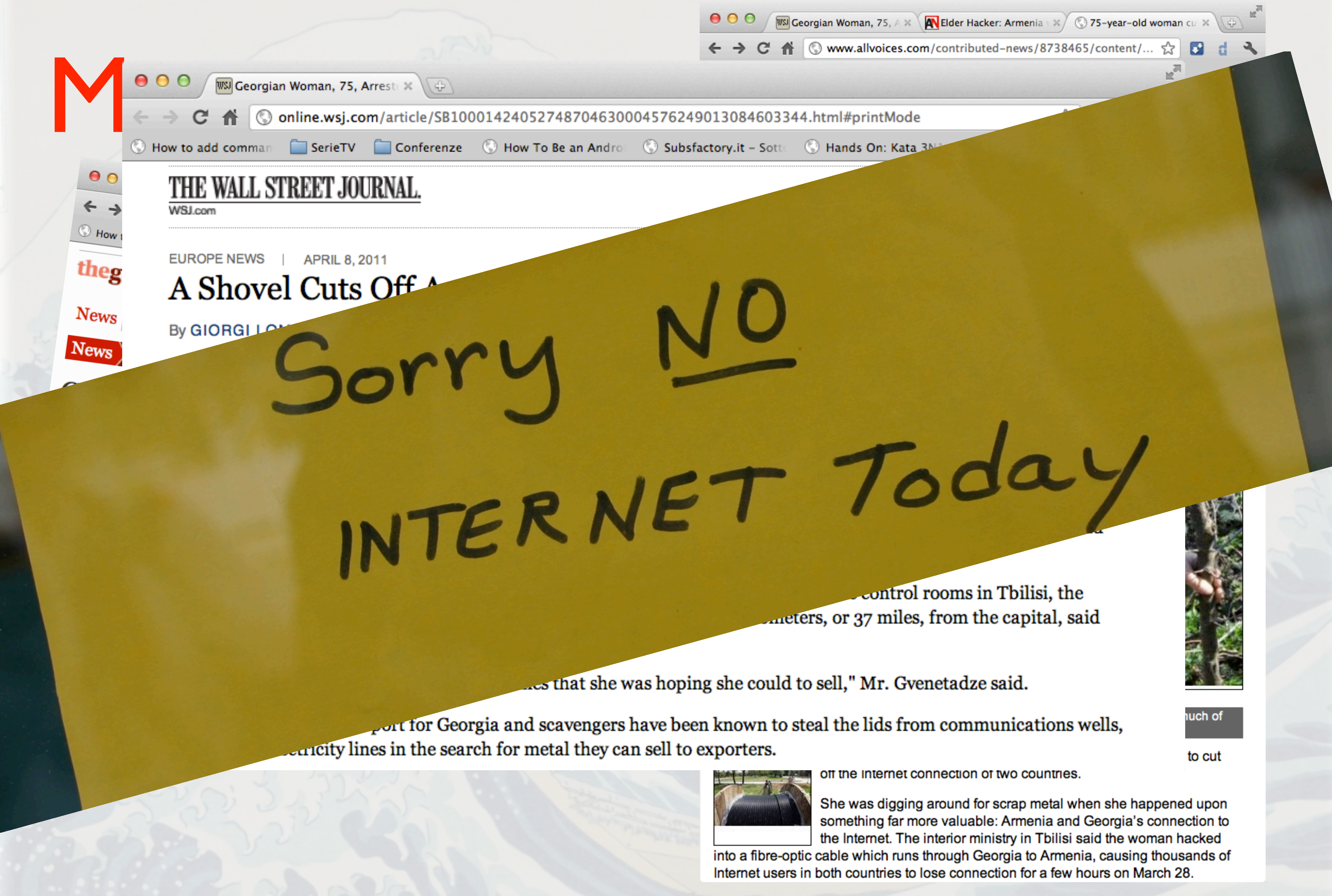
...uch of ... to cut off the internet connection of two countries.



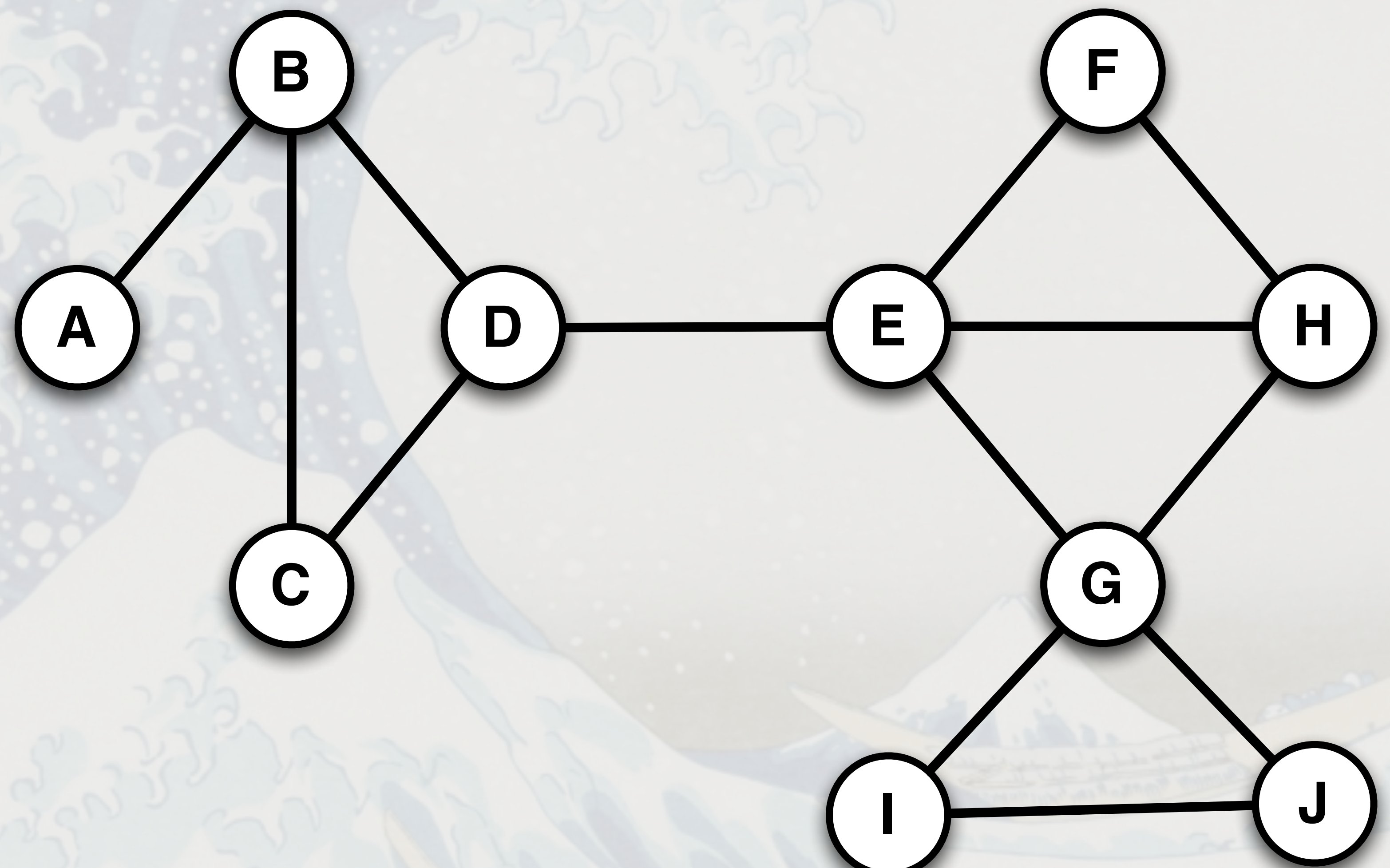
She was digging around for scrap metal when she happened upon something far more valuable: Armenia and Georgia's connection to the Internet. The interior ministry in Tbilisi said the woman hacked into a fibre-optic cable which runs through Georgia to Armenia, causing thousands of Internet users in both countries to lose connection for a few hours on March 28.

Geor... whole...
Entire co... slices thr...

Tom Parfitt in ...
guardian.co.uk

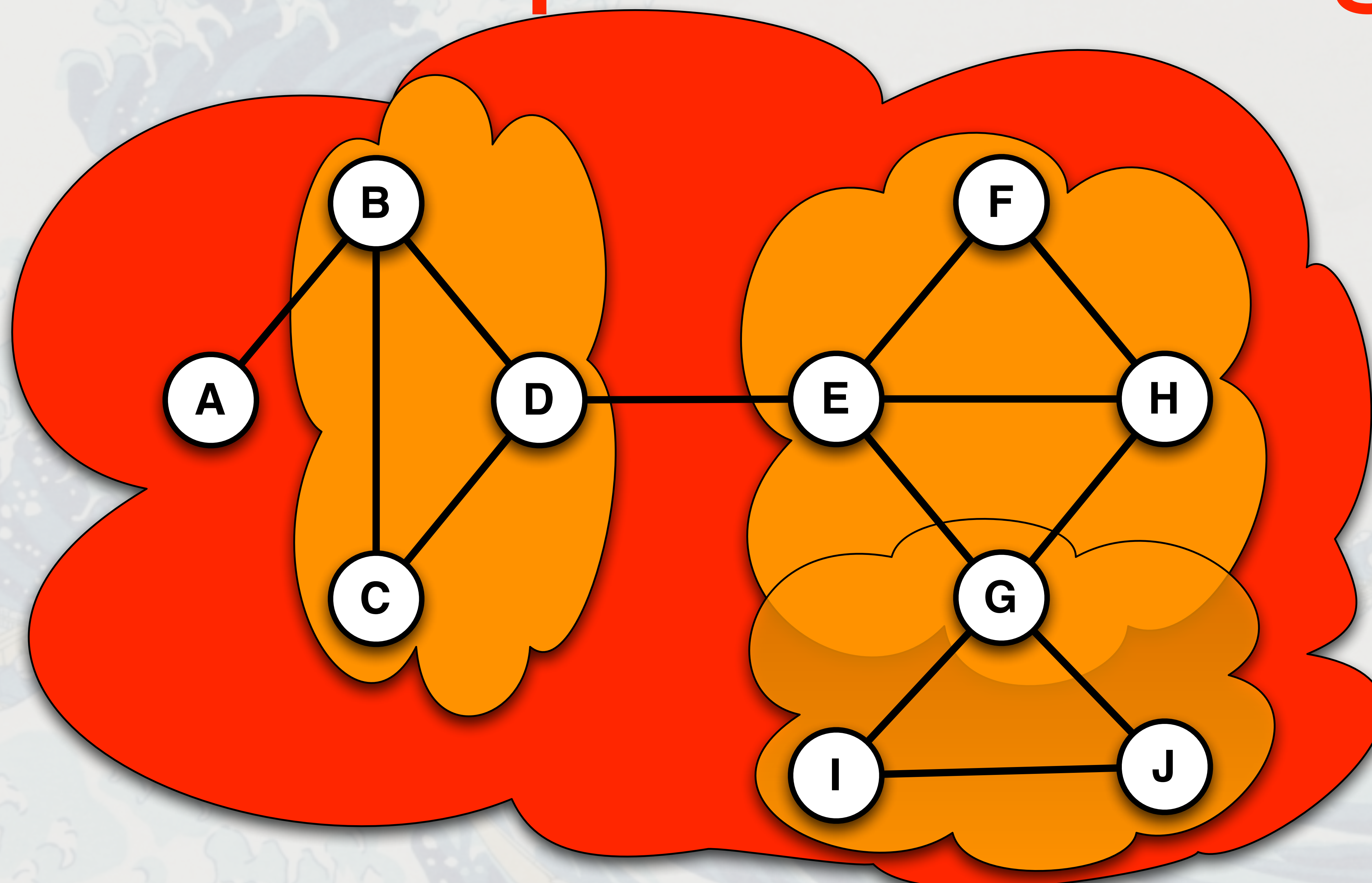

Georgian Woman, 75,
Elder Hacker: Armenia
75-year-old woman
www.allvoices.com/contributed-news/8738465/content/...
Georgian Woman, 75, Arrest
online.wsj.com/article/SB10001424052748704630004576249013084603344.html#printMode
How to add comman
SerieTV
Conferenze
How To Be an Andro
Subsfactory.it – Sott
Hands On: Kata
THE WALL STREET JOURNAL.
WSJ.com
EUROPE NEWS | APRIL 8, 2011
A Shovel Cuts Off A
By GIORGI L
Sorry NO
INTERNET Today
control rooms in Tbilisi, the
meters, or 37 miles, from the capital, said
that she was hoping she could to sell," Mr. Gvenetadze said.
port for Georgia and scavengers have been known to steal the lids from communications wells,
electricity lines in the search for metal they can sell to exporters.
off the internet connection of two countries.
She was digging around for scrap metal when she happened upon something far more valuable: Armenia and Georgia's connection to the Internet. The interior ministry in Tbilisi said the woman hacked into a fibre-optic cable which runs through Georgia to Armenia, causing thousands of Internet users in both countries to lose connection for a few hours on March 28.

Basic Graph Terminology



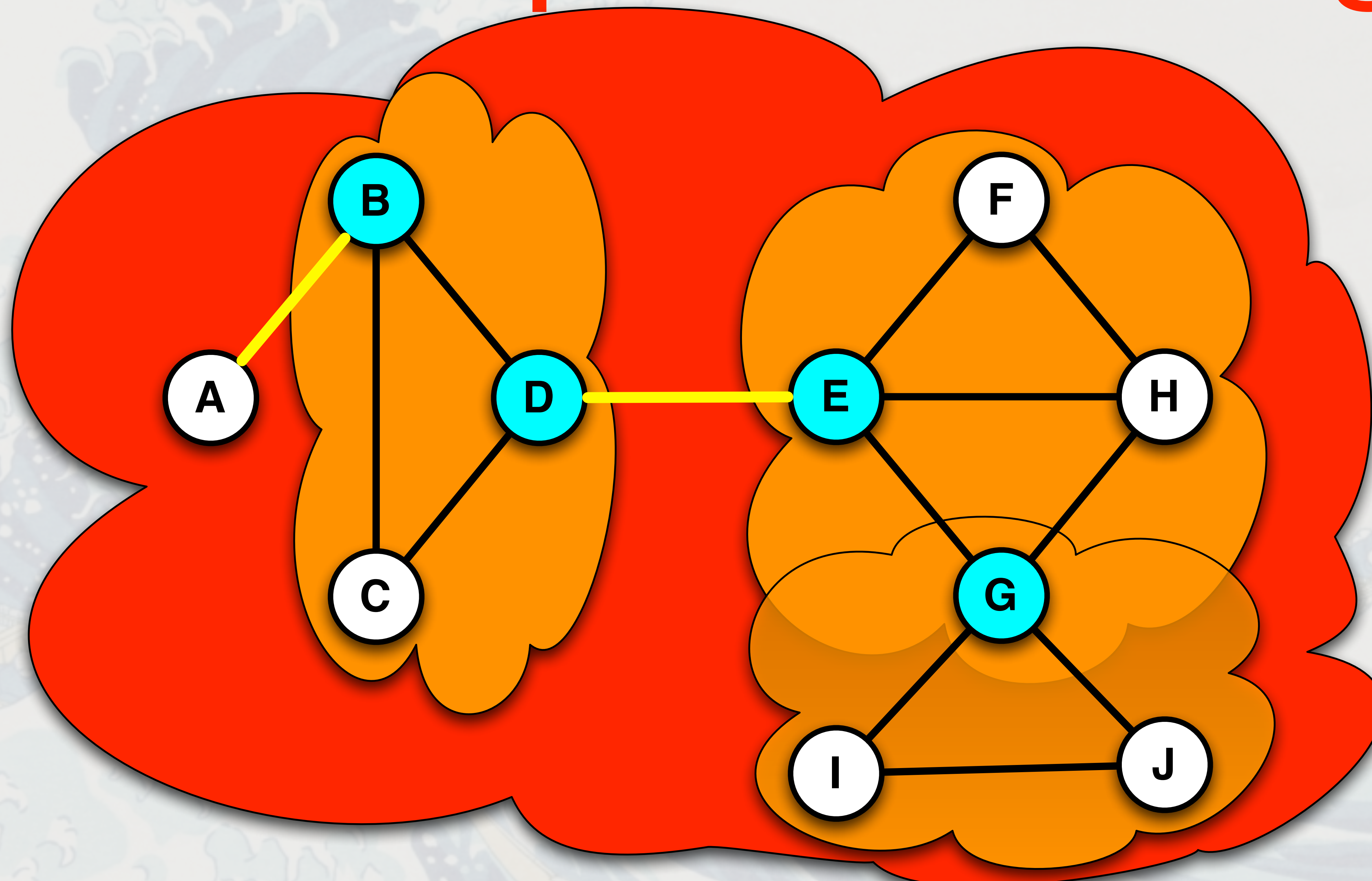
A simple connected undirected graph

Basic Graph Terminology



Connected and Biconnected Components

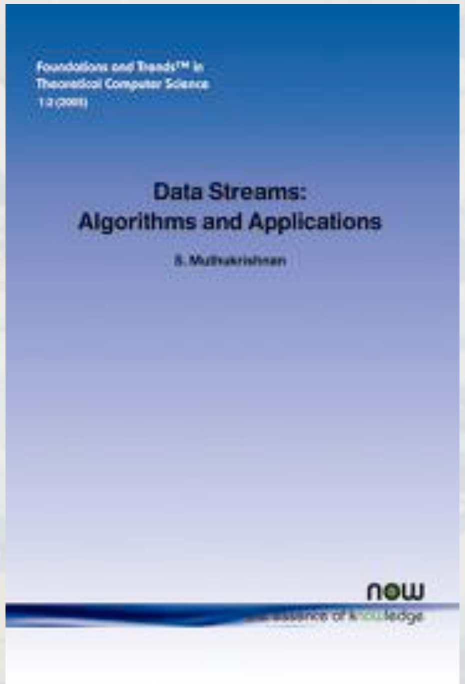
Basic Graph Terminology



Articulation Points and Bridges

“Space-aware” computational models

- External memory model (Aggarwal and Vitter 1998)
- Cache oblivious (Frigo *et al.* 1999) vs cache aware
- Several datastream models (see Muthu’s book!)



Datastream models

There are several datastream models that differs in various aspects:

- the number of passes (can be limited to one pass only!)
- the memory available (graph problems: full vs semi streaming)
- the ability to read/write additional streams

Our setting...

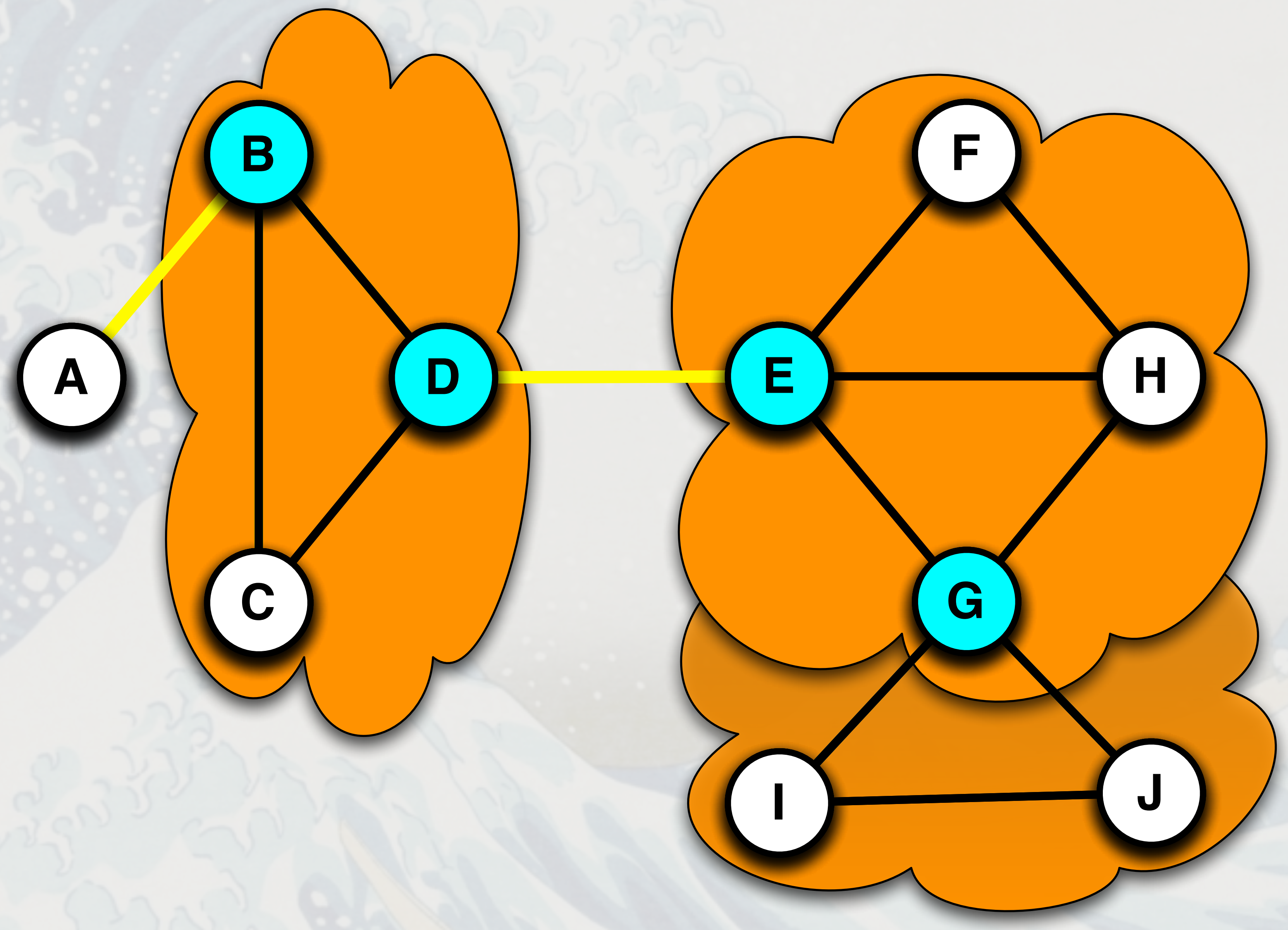
- Semi-streaming model (i.e., space=$O(n)$)
- Graph represented by a stream of edges in any (adversarial) order

The Navigational Sketch

Let us “represent” the graph in the following way, in a forest with two types of edges:

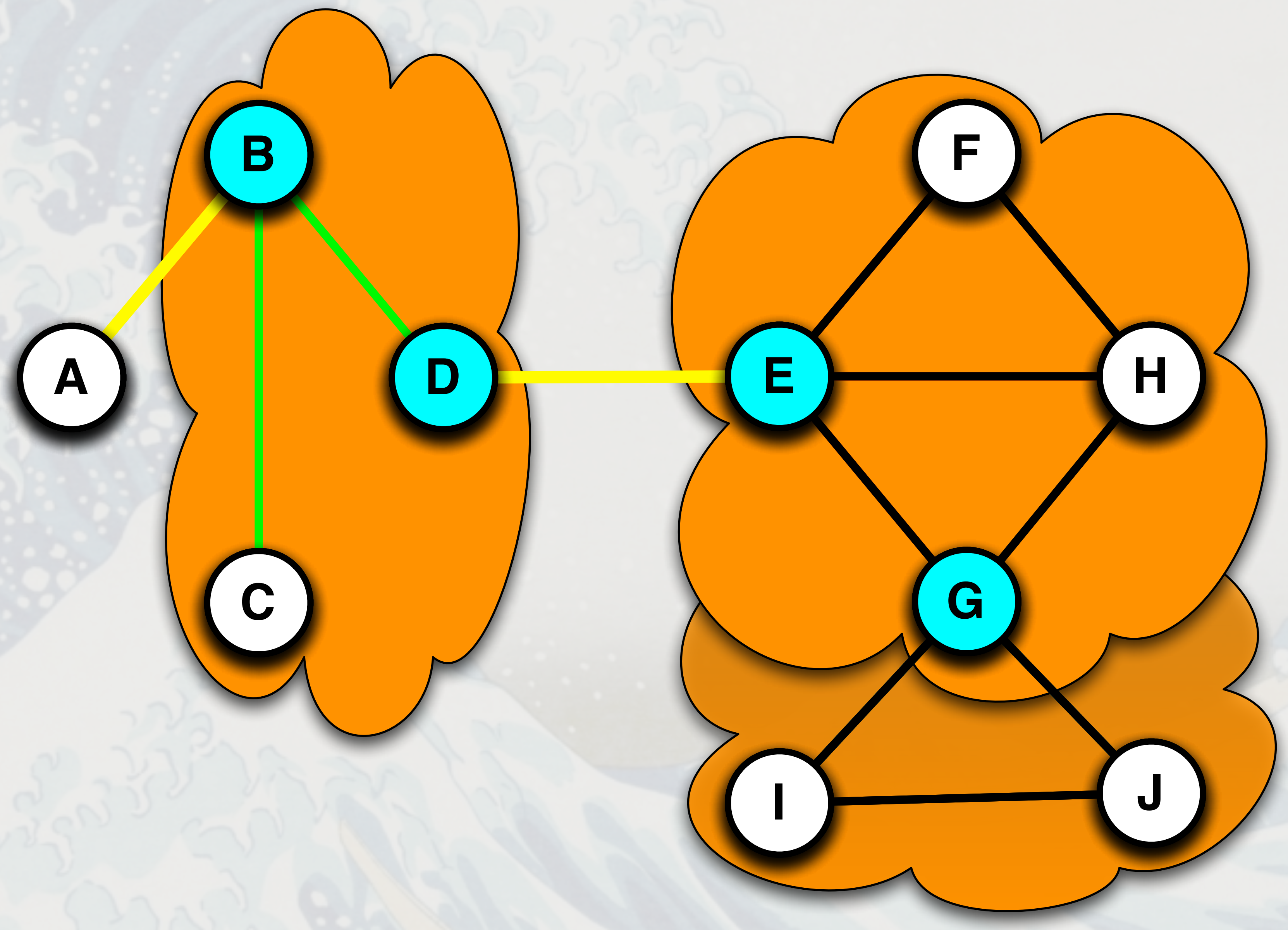
- each biconnected components is represented by a tree joined by COLORED edges (each color is unique for each BCC)
- each bridge is represented by a SOLID edge;
- it follows that each connected component is represented by a tree in the forest

The Navigational Sketch



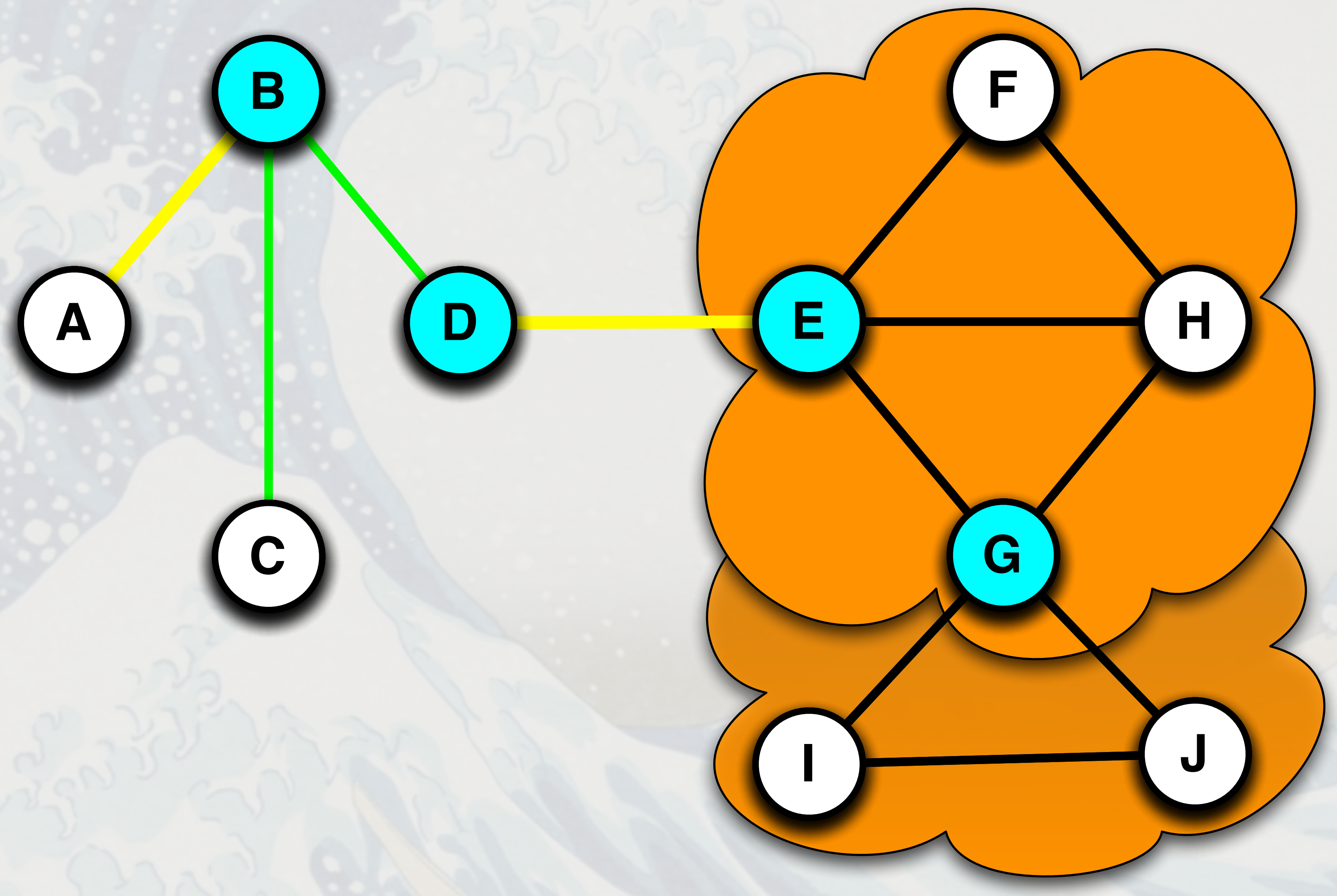
Solid edges, i.e. Bridges, are yellow.

The Navigational Sketch



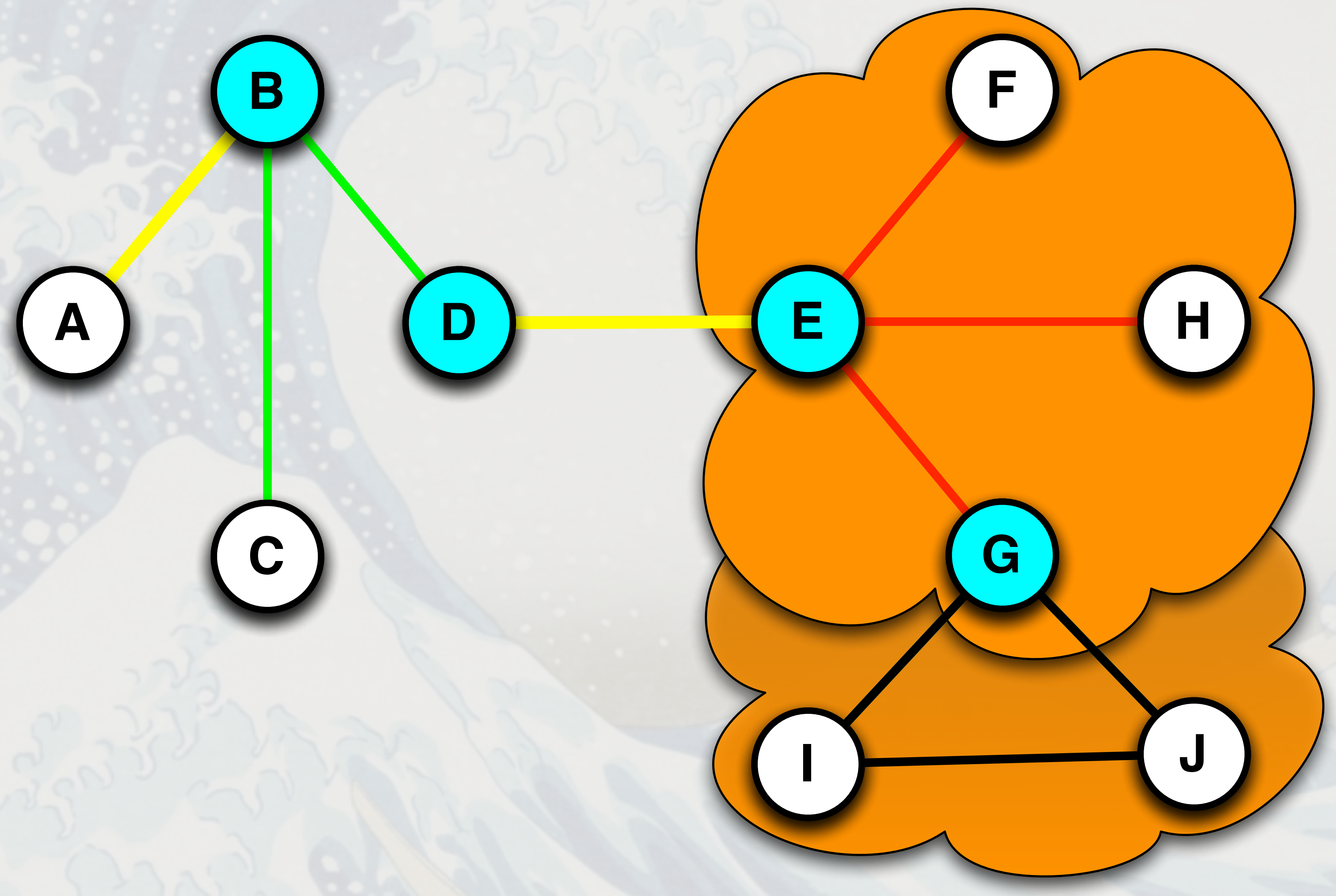
Solid edges, i.e. Bridges, are yellow.

The Navigational Sketch



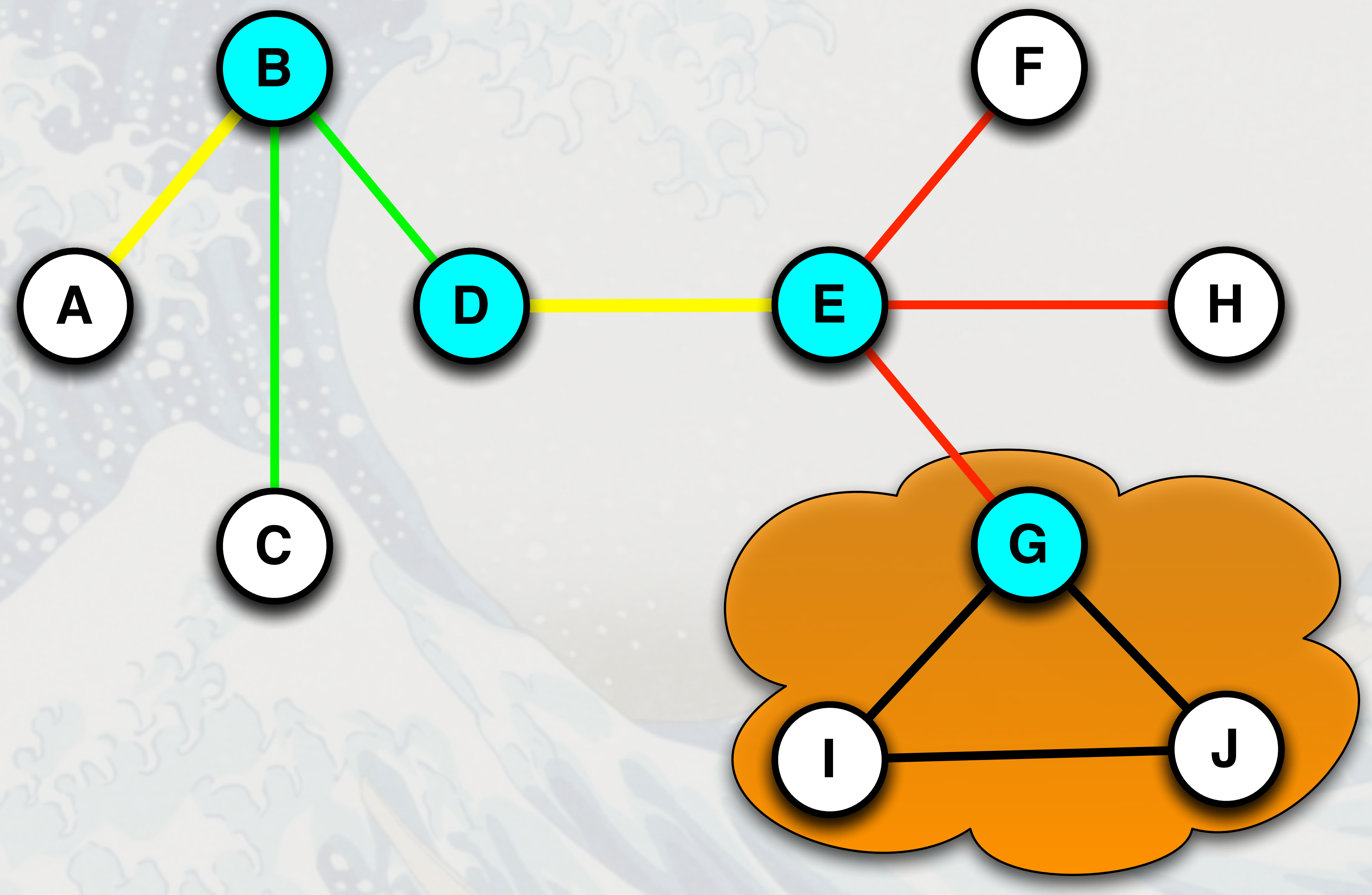
Solid edges, i.e. Bridges, are yellow.

The Navigational Sketch



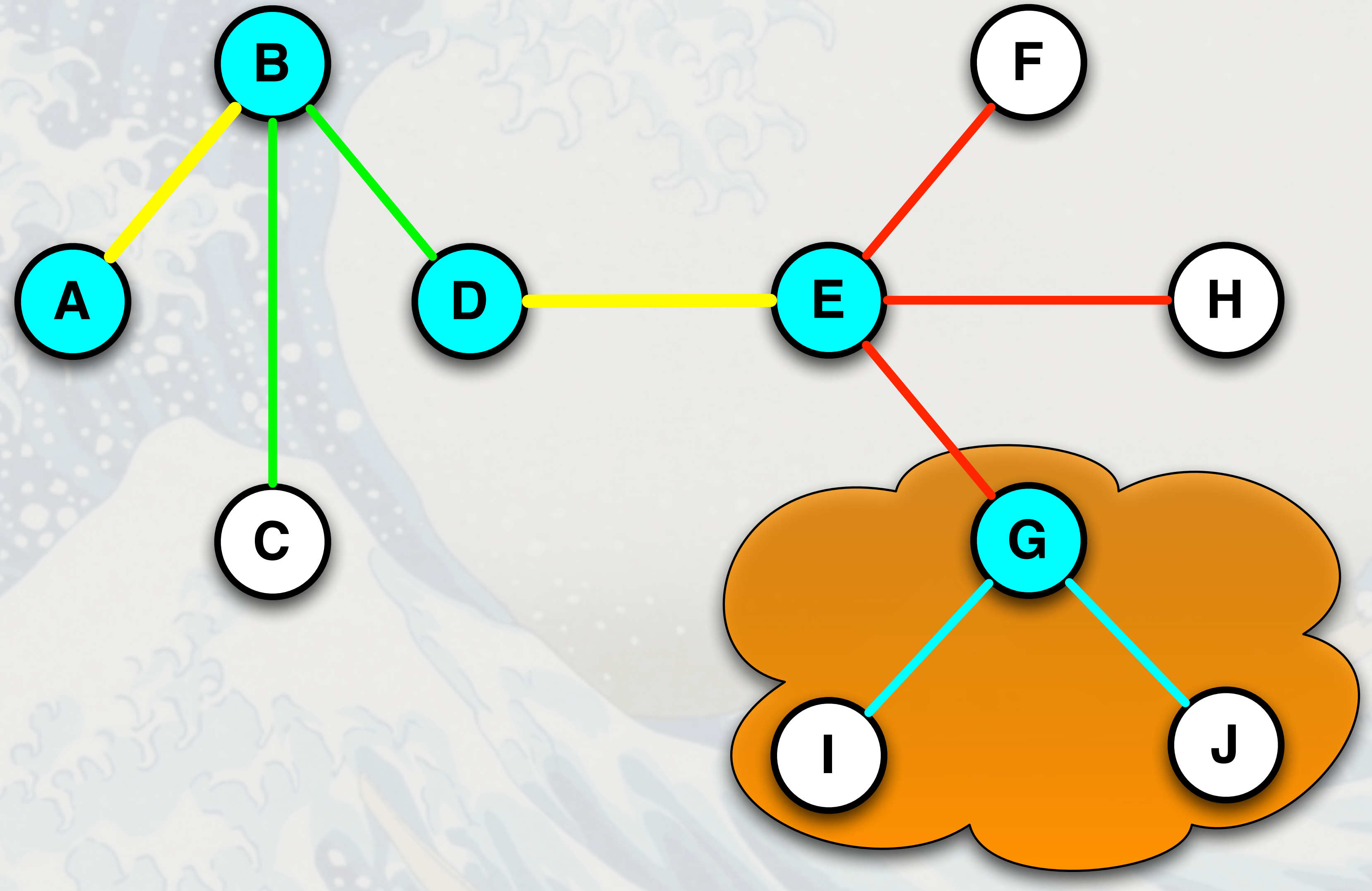
Solid edges, i.e. Bridges, are yellow.

The Navigational Sketch



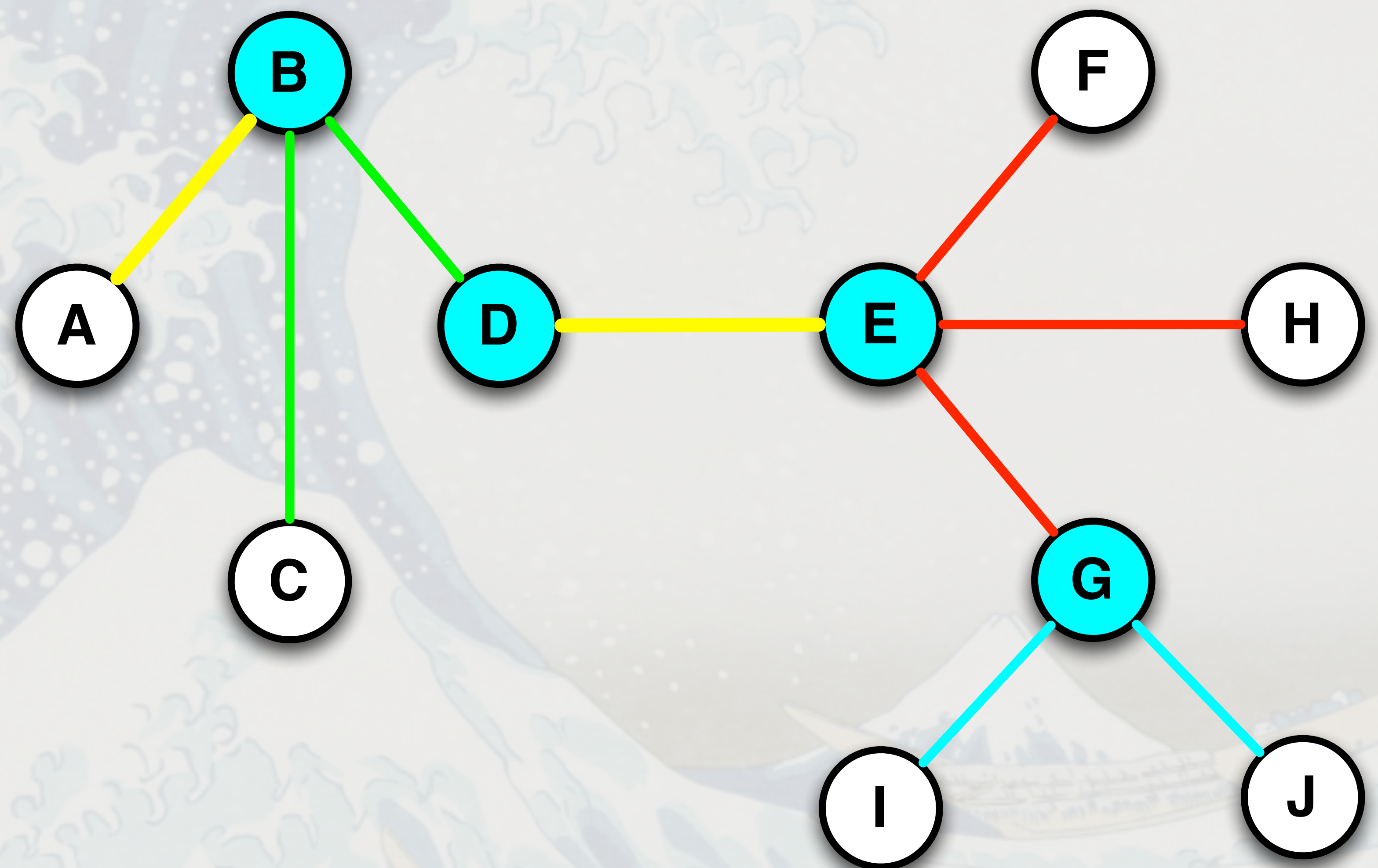
Solid edges, i.e. Bridges, are yellow.

The Navigational Sketch



Solid edges, i.e. Bridges, are yellow.

The Navigational Sketch



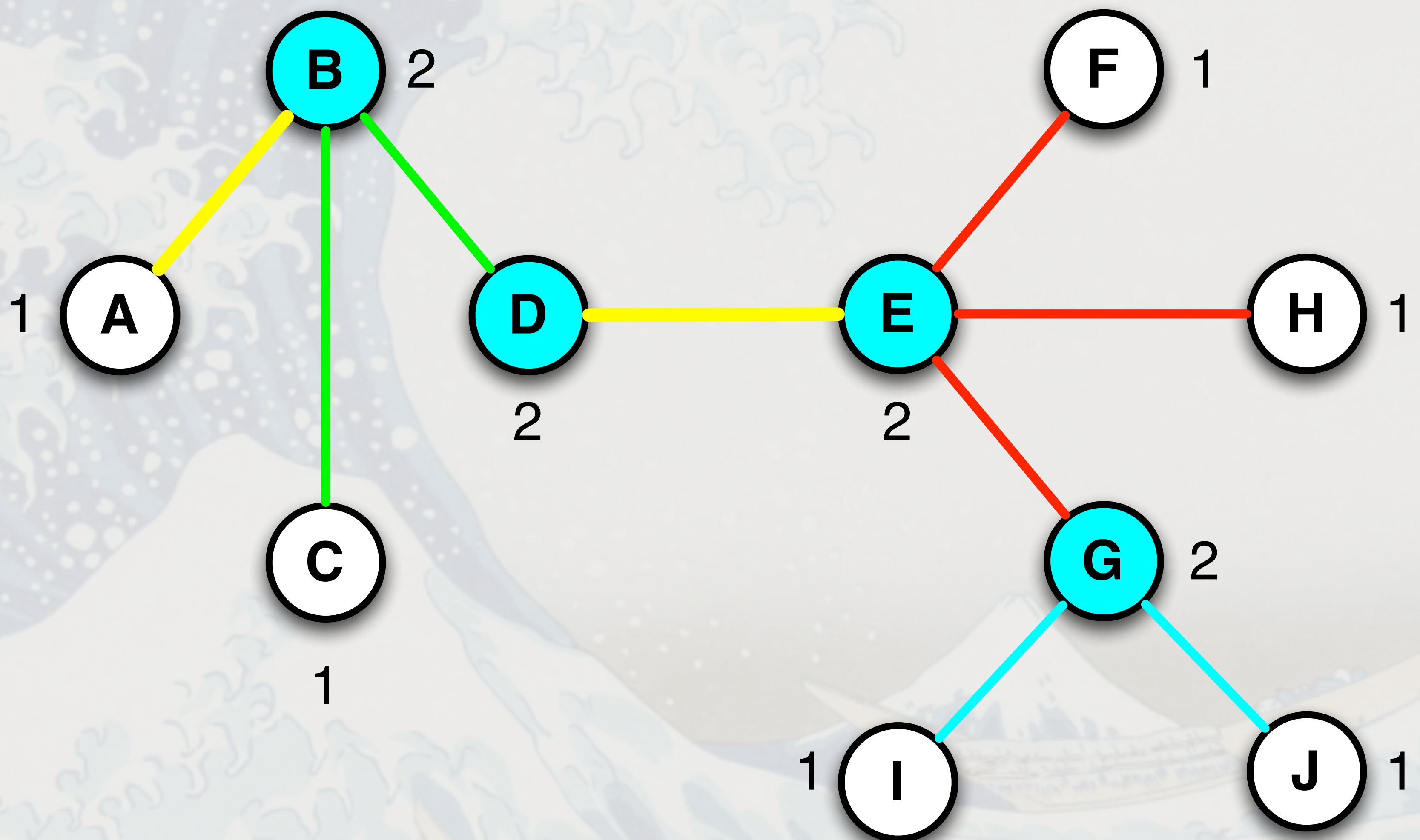
Solid edges, i.e. Bridges, are yellow.

Color degree of a node

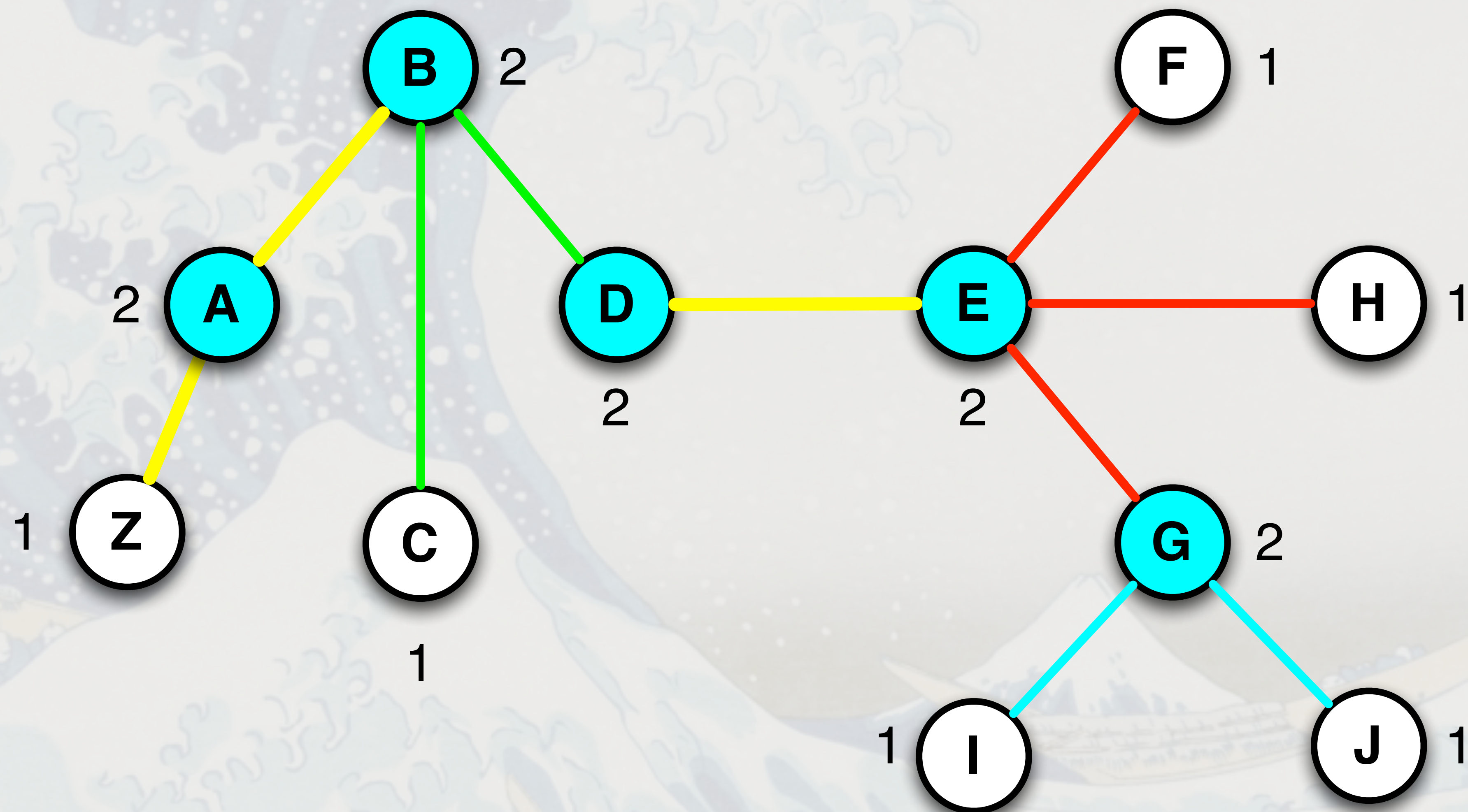
Let us define the **color degree** of a node *i* in a Navigational Sketch as the number of solid edges incident to *i* plus the number of distinct colors of edges incident to *i*.

It is easy to see that each node whose color degree is greater than one is an **articulation point** of the original graph.

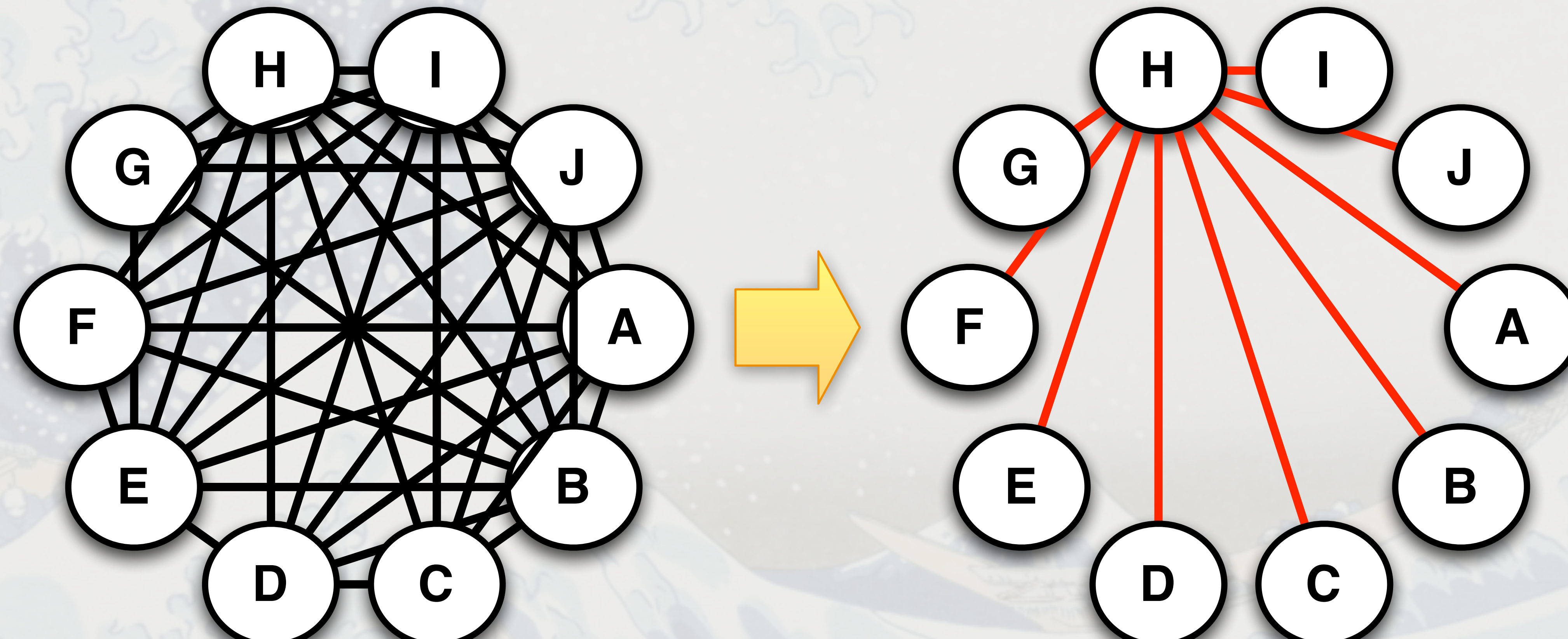
Color degree of a node



Color degree of a node (2)



Edges dropped...



The Algorithm: At First Look (AFL)

It mantains in main memory a Navigational Sketch (NS), built from scratch; at each step the algorithm looks at the current edge from the stream and, *at first look*, it decides on the corresponding action:

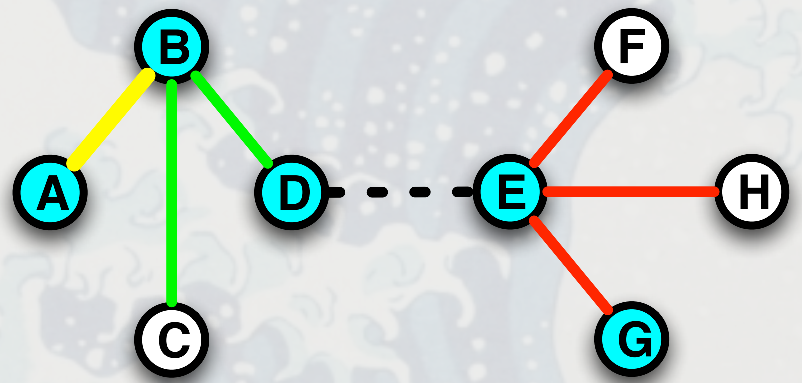
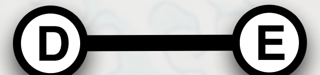
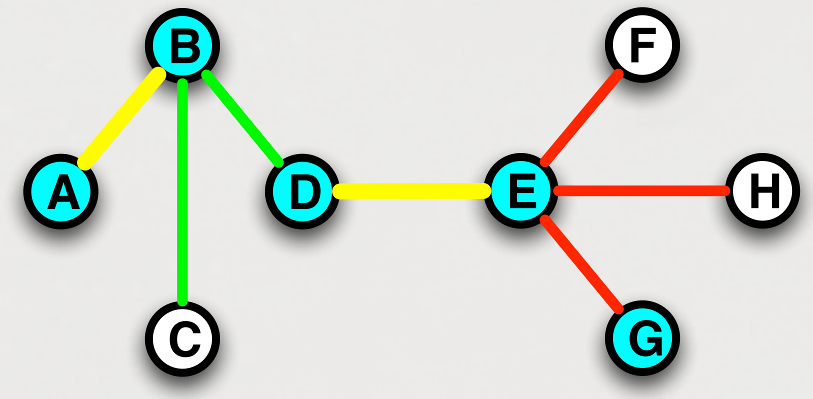
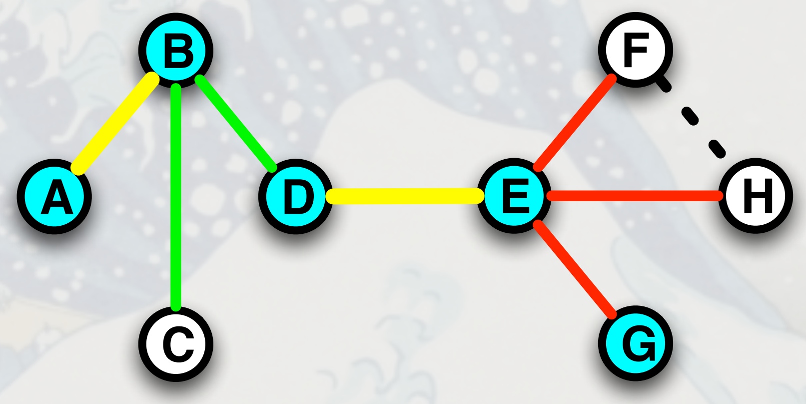
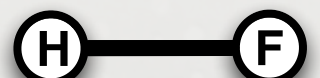
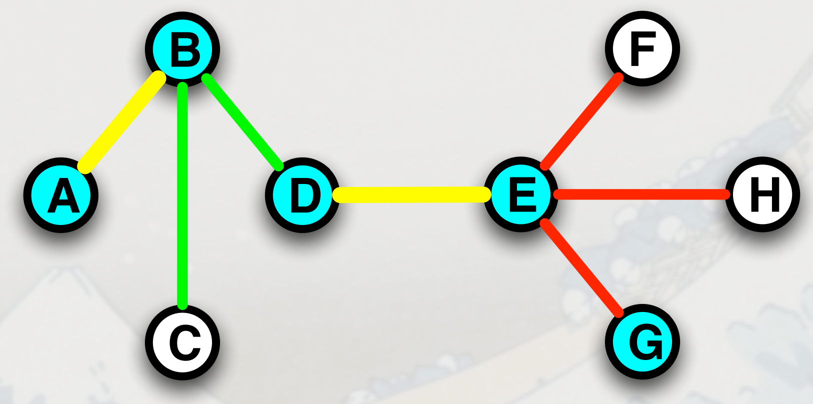
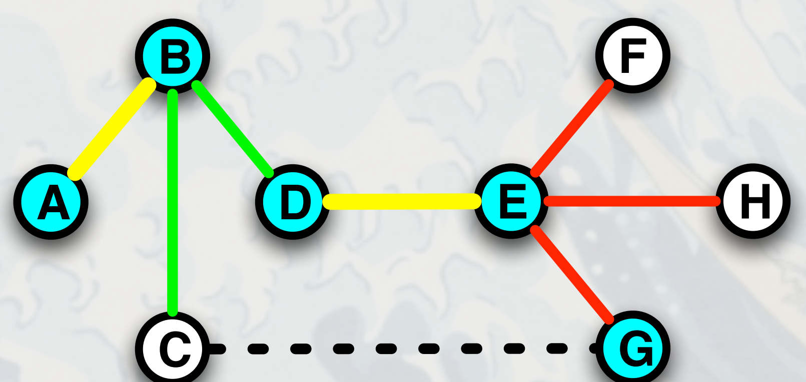
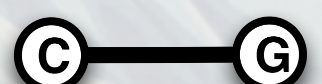
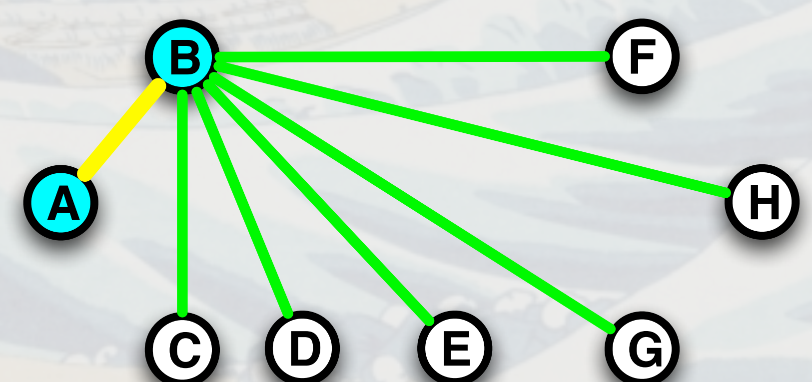
1. if the current edge joins two different trees
   -> it is added to the forest as a solid edge;

2. if the edge connects two nodes inside the same connected component, we need to distinguish two cases:

   2.1. all the edges in the path are of the same color
   -> the edge is dropped

   2.2. the path is made of different edge types
   -> we "rearrange" the NS

AFL cases

CASE	CURRENT NS	EDGE	UPDATED NS
1			
2.1			
2.2			

Theoretical results

Bounds of the algorithm:

- Space: $O(n \log n)$
- Worst Case Overall Processing Time: $O(m\alpha(m,n) + n \log n)$
- Per Item Processing Time (amortized): $O(\alpha(m,n) + (n/m) \log n) \approx O(\alpha(m,n))$ if the graph is dense enough (i.e. $m > n \log n$)

$\alpha(m,n)$ is a very slowly growing function, a functional inverse of the Ackermann's function, and for every "practical" value of m and n, it holds that $\alpha(m, n) \leq 4$.

Experimental Results

- We run our experiments on an off-the-shelf computer: a dual boot (Windows Vista and Ubuntu Linux) laptop Dell XPS M1330 (4Gb RAM, Intel Core2 Duo T8100 2.1GHz).
- We implemented AFL in the C programming language, using the gcc compiler under both Windows and Ubuntu.
- We used real world graph taken from:

#	Author	URL
1	A. Arenas	`http://deim.urv.cat/~aarenas/data/welcome.htm`
2	M. Newman	`http://www-personal.umich.edu/~mejn/netdata/`
3	Pajek datasets (V. Batagelj and A. Mrvar)	`http://vlado.fmf.uni-lj.si/pub/networks/data/`
4	Route Views	`http://www.routeviews.org/`
5	P. Boldi and S. Vigna [4]	`http://law.dsi.unimi.it/`

Experimental Results

Graph	Type	Data Source	Disk Space	Number of Nodes: $n = \|V\|$	Number of Edges: $m = \|E\|$	Average Degree: $\frac{m}{n}$	Density Factor: $\frac{n \log n}{m}$	Max # *touched* edges per operation	Avg. # *touched* edges per operation	Overall Processing Time (t)	Amortized PIPT: $\frac{t}{m}$	Edges processed per second: $\frac{m}{t}$
`Yeast`	biology	3	70 Kb	7.1k	2.4k	0,33	38,90	6	0,20	< 0.001	$\approx$ 1.00E-10	-
`DutchElite`	economy	3	54 Kb	4.7k	5.2k	1,10	11,11	8	1,04	< 0.001	$\approx$ 1.00E-10	-
`USpowerGrid`	technology	3	136 Kb	4.9k	9.9k	2,00	6,13	15	0,34	< 0.1	$\approx$ 9.60E-6	-
`email`	social	1	93 Kb	1.1k	11k	9,67	1,05	2	0,01	< 0.1	$\approx$ 5.56E-6	-
`AS`	Aut. Systems	4	670 Kb	57k	57k	0,88	18,24	4	0,77	< 0.1	$\approx$ 8.17E-7	-
`PairsP`	similarity	3	737 Kb	10.6k	72k	6,78	1,97	7	0,09	< 0.1	$\approx$ 4.30E-7	-
`dic-28`	linguistic	3	1 Mb	52.6k	89k	1,69	9,28	9	0,64	< 0.1	$\approx$ 5.17E-7	-
`foldoc`	linguistic	3	1.3 Mb	13.3k	119.8k	8,97	1,53	6	0,11	< 0.1	$\approx$ 5.25E-7	-
`wordnet3`	linguistic	3	1.6 Mb	82.6k	125.5k	1,52	10,76	8	0,28	< 0.2	$\approx$ 9.87E-7	-
`eatRS`	linguistic	3	3.8 Mb	23.2k	325.5k	14,02	1,03	9	0,06	< 0.2	$\approx$ 3.87E-7	-
`hep-th-new`	citation	2	4.2 Mb	27.7k	352.7k	12,70	1,16	7	0,07	< 0.3	$\approx$ 6.20E-7	-
`cnr-2000`	web	5	44.7 Mb	325k	3.2M	9,88	1,85	10	0,06	< 3	$\approx$ 1.02E-6	$\approx$ 1M
`eu-2005`	web	5	270.8 Mb	862k	19.2M	22,3	0,88	7	0,04	< 20	$\approx$ 1.06E-6	$\approx$ 1M
`indochina-2004`	web	5	3 Gb	7.4M	194.1M	26,18	0,87	72	0,03	< 200	$\approx$ 1.02E-6	$\approx$ 1M
`uk-2002`	web	5	5 Gb	18.5M	298.1M	16,1	1,50	91	0,05	< 300	$\approx$ 1.01E-6	$\approx$ 1M
`it-2004`	web	5	20.5 Gb	41.2M	1.1G	27,87	0,91	255	0,07	< 600	$\approx$ 5.80E-7	$\approx$ 2M

Graph	Type	Data Source	Disk Space	Number of Nodes: $n = \|V\|$	Number of Edges: $m = \|E\|$	Average Degree: $\frac{m}{n}$	Density Factor: $\frac{n \log n}{m}$	Max # *touched* edges per opera	Avg. # *touched* edges per opera
`Yeast`	biology	3	70 Kb	7.1k	2.4k	0,33	38,90	6	0,20
`DutchElite`	economy	3	54 Kb	4.7k	5.2k	1,10	11,11	8	1,04
`USpowerGrid`	technology	3	136 Kb	4.9k	9.9k	2,00	6,13	15	0,34
`email`	social	1	93 Kb	1.1k	11k	9,67	1,05	2	0,01
`AS`	Aut. Systems	4	670 Kb	57k	57k	0,88	18,24	4	0,77
`PairsP`	similarity	3	737 Kb	10.6k	72k	6,78	1,97	7	0,09
`dic-28`	linguistic	3	1 Mb	52.6k	89k	1,69	9,28	9	0,64
`foldoc`	linguistic	3	1.3 Mb	13.3k	119.8k	8,97	1,53	6	0,11
`wordnet3`	linguistic	3	1.6 Mb	82.6k	125.5k	1,52	10,76	8	0,28
`eatRS`	linguistic	3	3.8 Mb	23.2k	325.5k	14,02	1,03	9	0,06
`hep-th-new`	citation	2	4.2 Mb	27.7k	352.7k	12,70	1,16	7	0,07
`cnr-2000`	web	5	44.7 Mb	325k	3.2M	9,88	1,85	10	0,06
`eu-2005`	web	5	270.8 Mb	862k	19.2M	22,3	0,88	7	0,04
`indochina-2004`	web	5	3 Gb	7.4M	194.1M	26,18	0,87	72	0,03
`uk-2002`	web	5	5 Gb	18.5M	298.1M	16,1	1,50	91	0,05
`it-2004`	web	5	20.5 Gb	41.2M	1.1G	27,87	0,91	255	0,07

	Data Source	Disk Space	Number of Nodes: $n = \|V\|$	Number of Edges: $m = \|E\|$	Average Degree: $\frac{m}{n}$	Density Factor: $\frac{n \log n}{m}$	Max # *touched* edges per opera	Avg. # *touched* edges per oper	Overall Processing Time (t)	Amortized PIPT: $\frac{t}{m}$	Edges processed per second: $\frac{m}{t}$
	3	70 Kb	7.1k	2.4k	0,33	38,90	6	0,20	< 0.001	$\approx$ 1.00E-10	-
	3	54 Kb	4.7k	5.2k	1,10	11,11	8	1,04	< 0.001	$\approx$ 1.00E-10	-
y	3	136 Kb	4.9k	9.9k	2,00	6,13	15	0,34	< 0.1	$\approx$ 9.60E-6	-
	1	93 Kb	1.1k	11k	9,67	1,05	2	0,01	< 0.1	$\approx$ 5.56E-6	-
ems	4	670 Kb	57k	57k	0,88	18,24	4	0,77	< 0.1	$\approx$ 8.17E-7	-
	3	737 Kb	10.6k	72k	6,78	1,97	7	0,09	< 0.1	$\approx$ 4.30E-7	-
	3	1 Mb	52.6k	89k	1,69	9,28	9	0,64	< 0.1	$\approx$ 5.17E-7	-
	3	1.3 Mb	13.3k	119.8k	8,97	1,53	6	0,11	< 0.1	$\approx$ 5.25E-7	-
	3	1.6 Mb	82.6k	125.5k	1,52	10,76	8	0,28	< 0.2	$\approx$ 9.87E-7	-
	3	3.8 Mb	23.2k	325.5k	14,02	1,03	9	0,06	< 0.2	$\approx$ 3.87E-7	-
	2	4.2 Mb	27.7k	352.7k	12,70	1,16	7	0,07	< 0.3	$\approx$ 6.20E-7	-
	5	44.7 Mb	325k	3.2M	9,88	1,85	10	0,06	< 3	$\approx$ 1.02E-6	$\approx$ 1M
	5	270.8 Mb	862k	19.2M	22,3	0,88	7	0,04	< 20	$\approx$ 1.06E-6	$\approx$ 1M
	5	3 Gb	7.4M	194.1M	26,18	0,87	72	0,03	< 200	$\approx$ 1.02E-6	$\approx$ 1M
	5	5 Gb	18.5M	298.1M	16,1	1,50	91	0,05	< 300	$\approx$ 1.01E-6	$\approx$ 1M
	5	20.5 Gb	41.2M	1.1G	27,87	0,91	255	0,07	< 600	$\approx$ 5.80E-7	$\approx$ 2M

What about internet connectivity?

www.guardian.co.uk/world/2011/apr/06/georgian-woman-cuts-web-acc...

theguardian

News | Sport | Comment | Culture | Business | Money | Life & style | Travel | Environme

News > World news > Georgia

Georgian woman cuts off web access to whole of Armenia

Entire country loses internet for five h... slices through cable whi...

Tweet this

Tom Pa... guardian...

online.wsj.com/article/SB10001424052748704630004576249013084603344.html#printMode

THE WALL STREET JOURNAL.
WSJ.com

EUROPE NEWS | APRIL 8, 2011

A Shovel Cuts Off Armenia's Internet

By GIORGI LOMSADZE

TBILISI—Georgia has arrested a 75-year-old woman who, with her shovel, left all of Armenia without access to the Internet for half a day, according to Georgian police.

Aishtan Shakarian, who lives in a small village near Georgian capital Tbilisi, was scavenging for scrap metal on March 28 when she damaged an international fiber-optic backbone cable that connects much of the southern Caucasus to Europe, a police spokesman said.

The damage sent almost all of Armenia offline for about 12 hours. It also triggered a partial shutdown in Georgia and service interruptions in Azerbaijan, according to Georgian police and telecommunications companies.

The damage Ms. Shakarian caused to the line with her shovel sent alarm signals to the control rooms in Tbilisi, the operators said. She was caught digging in the village of Ksani, some 60 kilometers, or 37 miles, from the capital, said police spokesman Zurab Gvenetadze.

"The woman was hunting for some copper lines that she was hoping she could to sell," Mr. Gvenetadze said.

Scrap metal is a staple export for Georgia and scavengers have been known to steal the lids from communications wells, and to strip electricity lines in the search for metal they can sell to exporters.

www.allvoices.com/contributed-news/8738465/content/...

Contributor Report | News Stories: 2 | Blog Posts: 19 | Videos: 0

75-year-old woman cuts off Internet to Georgia and Armenia

Tbilisi : Georgia | Apr 10, 2011 at 1:39 AM PDT

BY mifha

VIEWS: 234

BACK | 3 of 4 | NEXT

Hayastan Shakarian denies cutting the fibre-optic cable that provides Internet for much of Armenia

A retired 75-year-old woman has single-handedly managed to cut off the Internet connection of two countries.

She was digging around for scrap metal when she happened upon something far more valuable: Armenia and Georgia's connection to the Internet. The interior ministry in Tbilisi said the woman hacked into a fibre-optic cable which runs through Georgia to Armenia, causing thousands of Internet users in both countries to lose connection for a few hours on March 28.

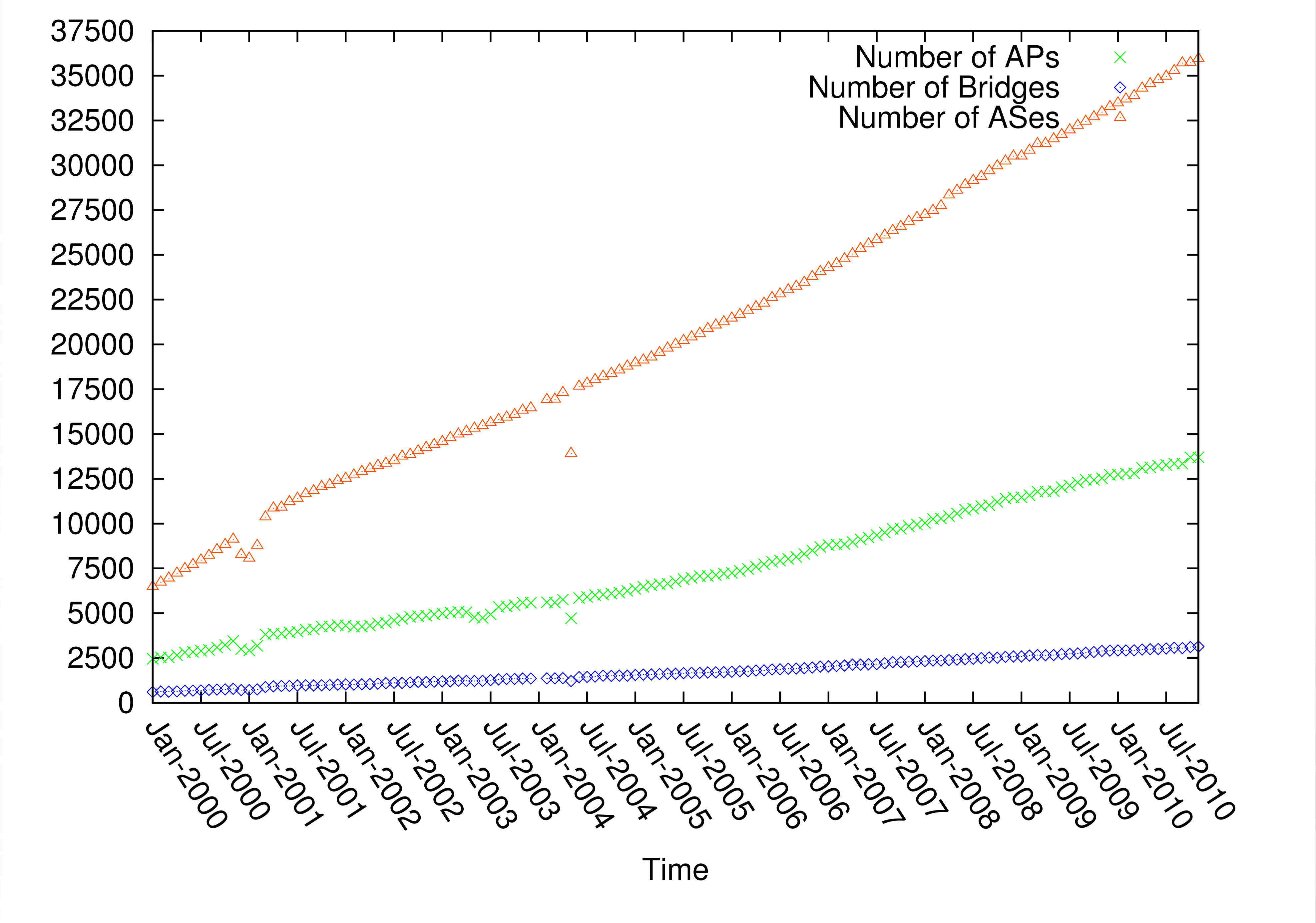
Autonomous Systems' historical data analysys

We implemented AFL in a tool, able to retrieve AS network data from the Oregon Route Views Project:

- historical data;
- real-time information (BGP announces)

We analyzed 10 years of AS networks: one sample per month, from Jan 2000 to Dec 2010

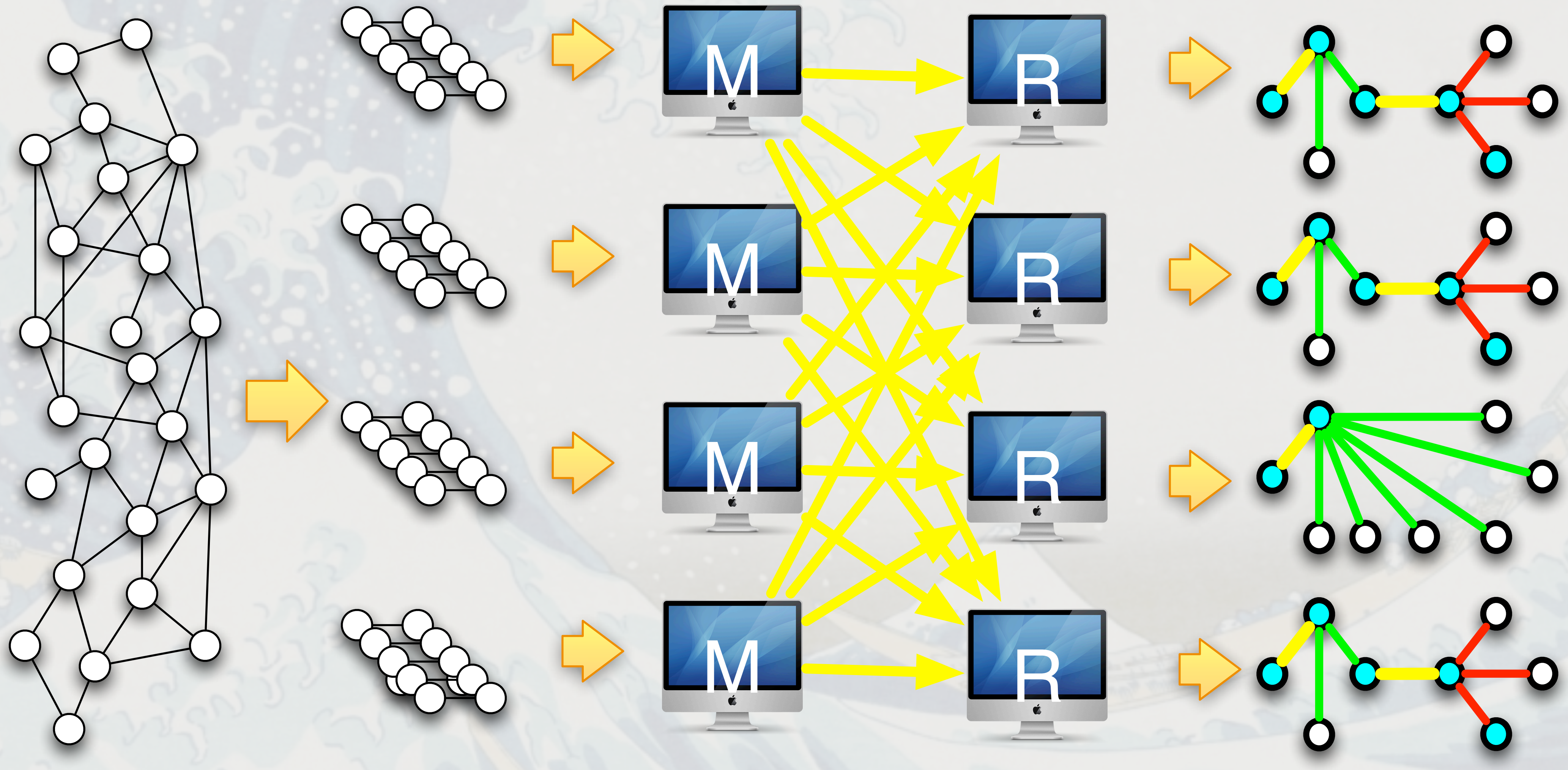
Articulation Points and Bridges



Biconnectivity in MapReduce

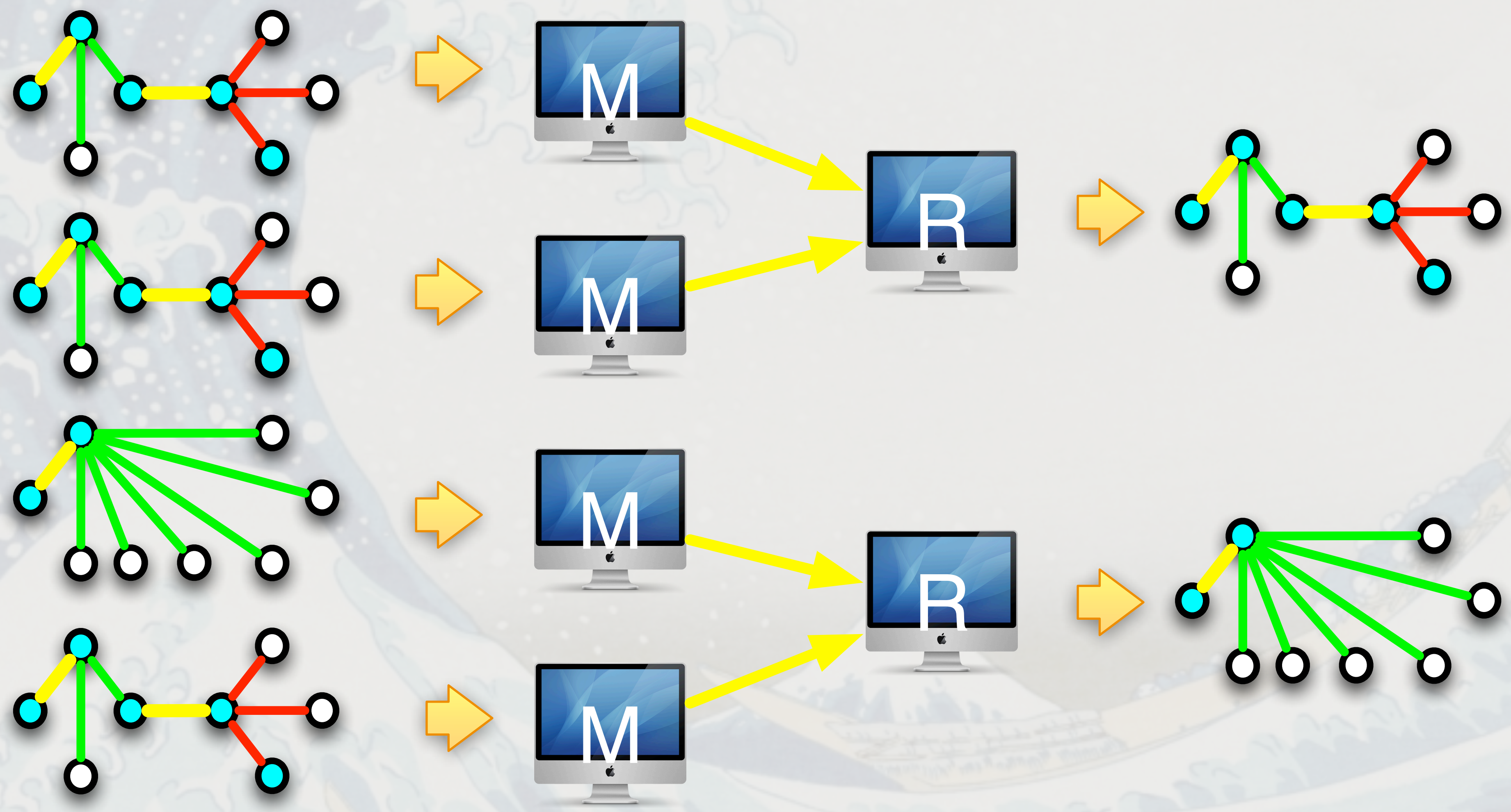
- We have seen, in Donatella's talk, the quest for an external (simple) connectivity algorithm...
- ... here, we focus on a semi-external approach to compute biconnected components (and APs, and Bridges, and CCs)...
- ... our approach is based (inspired!) on Navigational Sketches!

A first approach...



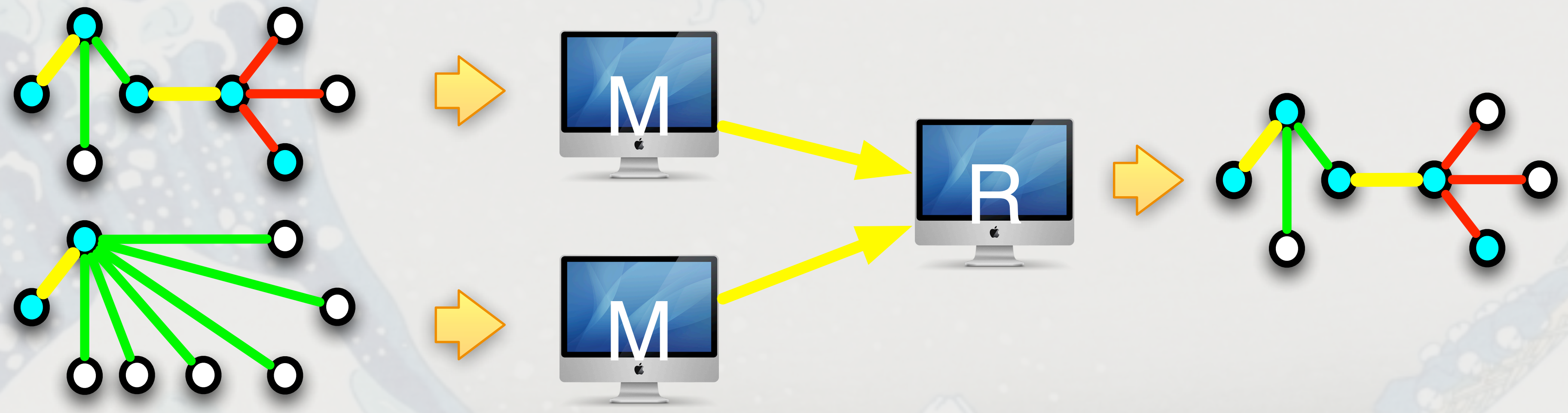
First round

A first approach...



Second round

A first approach...

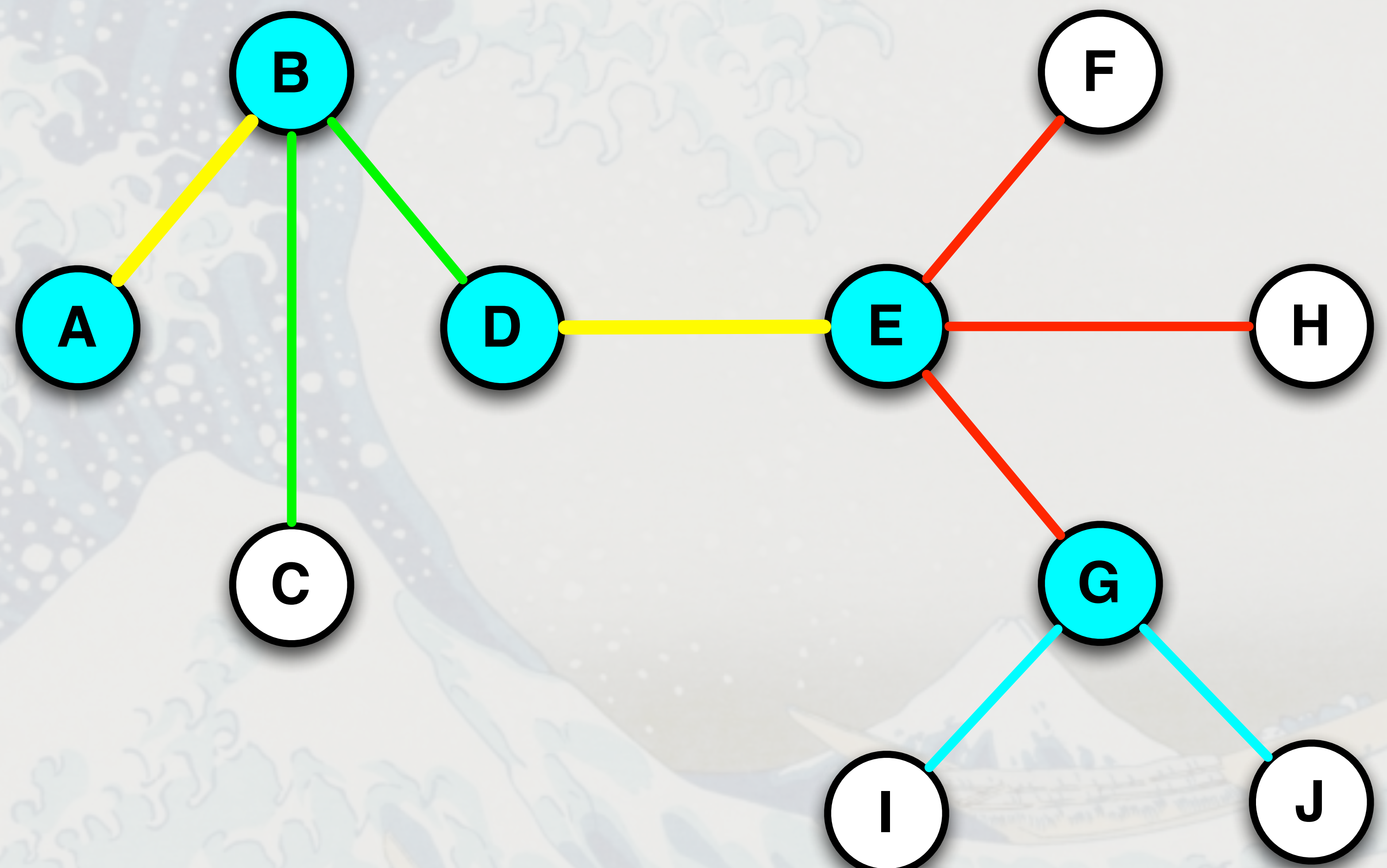


Third and final round

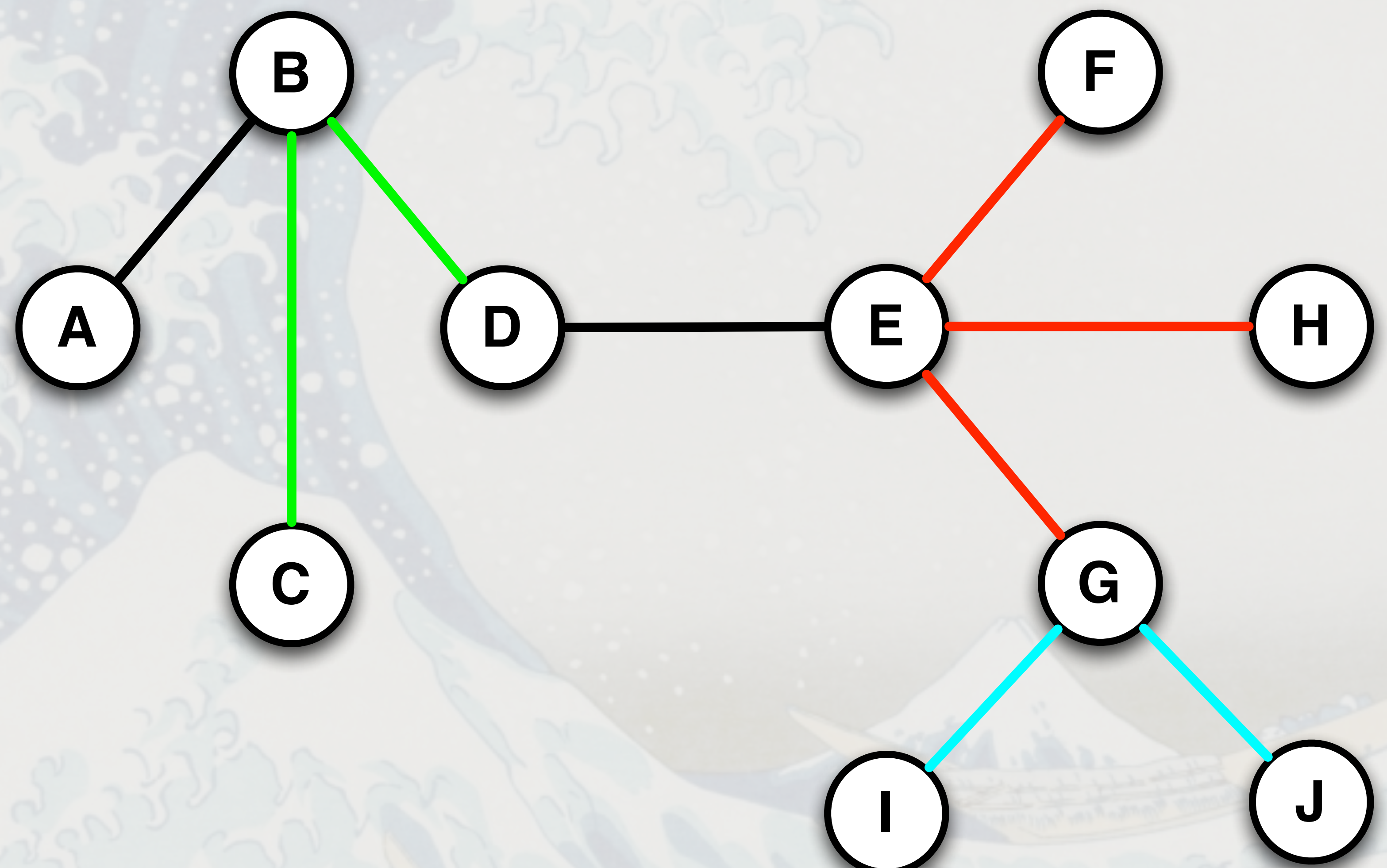
...but it has few drawbacks

- We cannot split a NS...
- ... therefore, each NS must move as a whole...
- ...we do not have the same freedom as with edges...

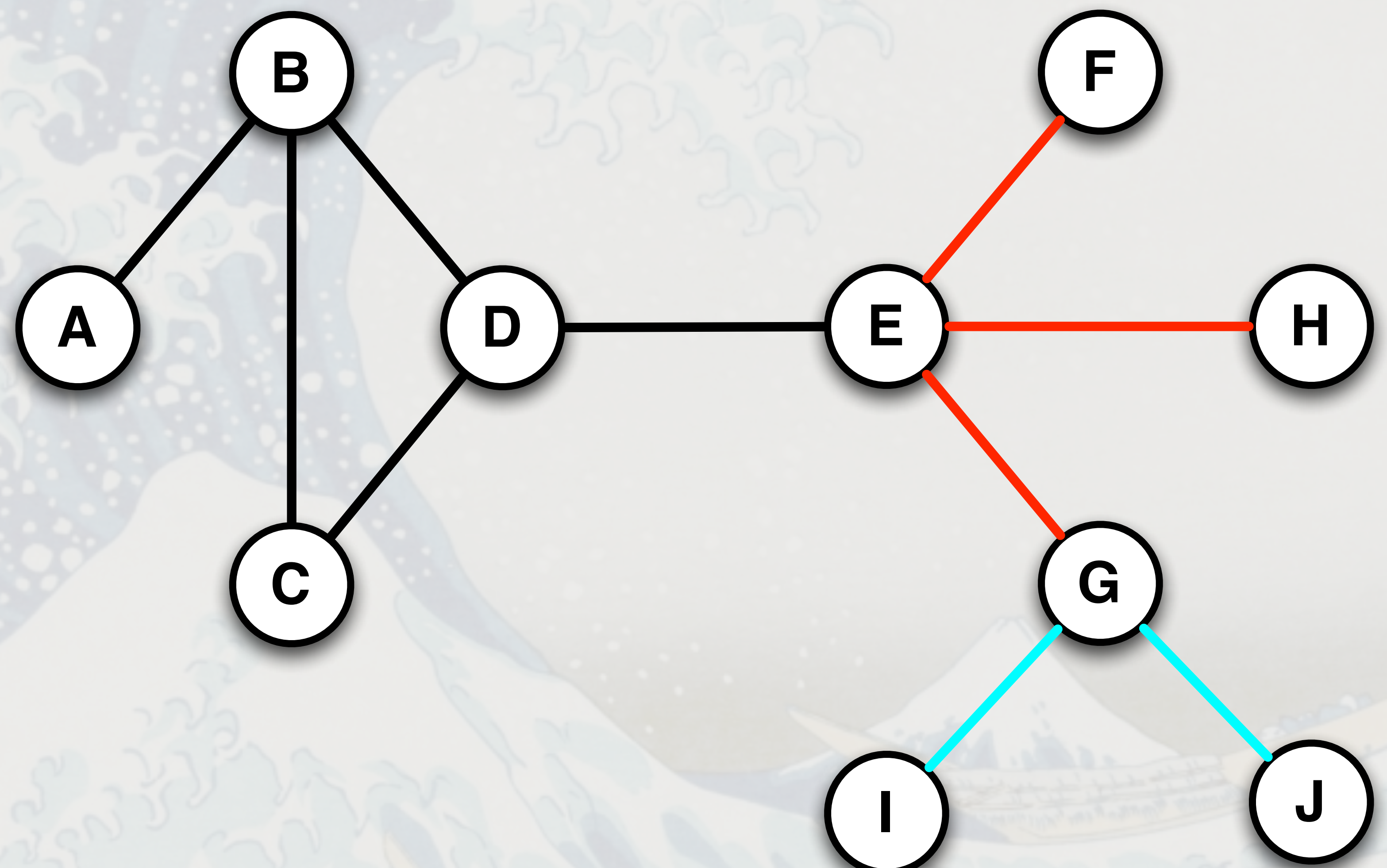
Improvement: from NS to a Minimal Equivalent Graph (MEG)



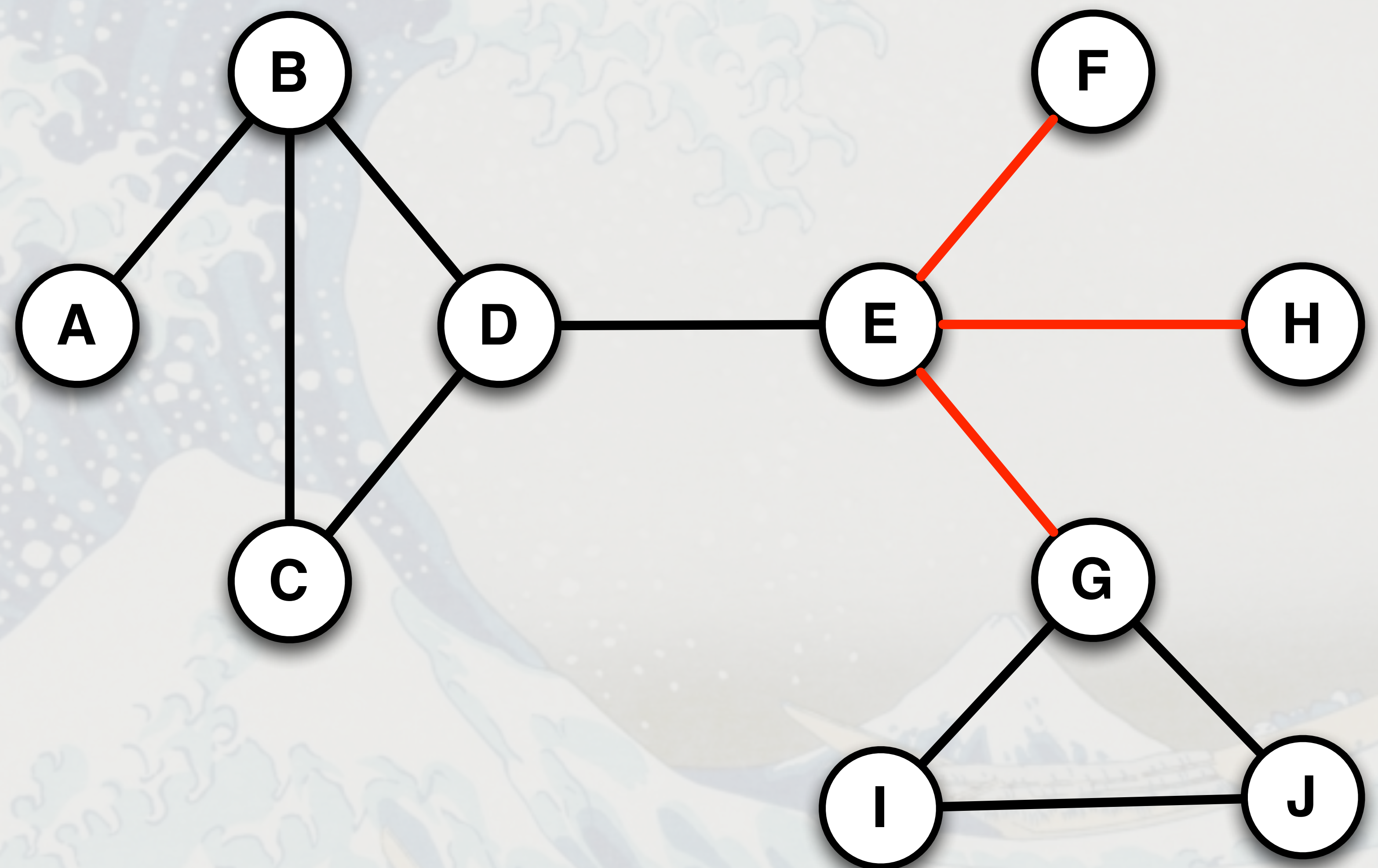
Improvement: from NS to a Minimal Equivalent Graph (MEG)



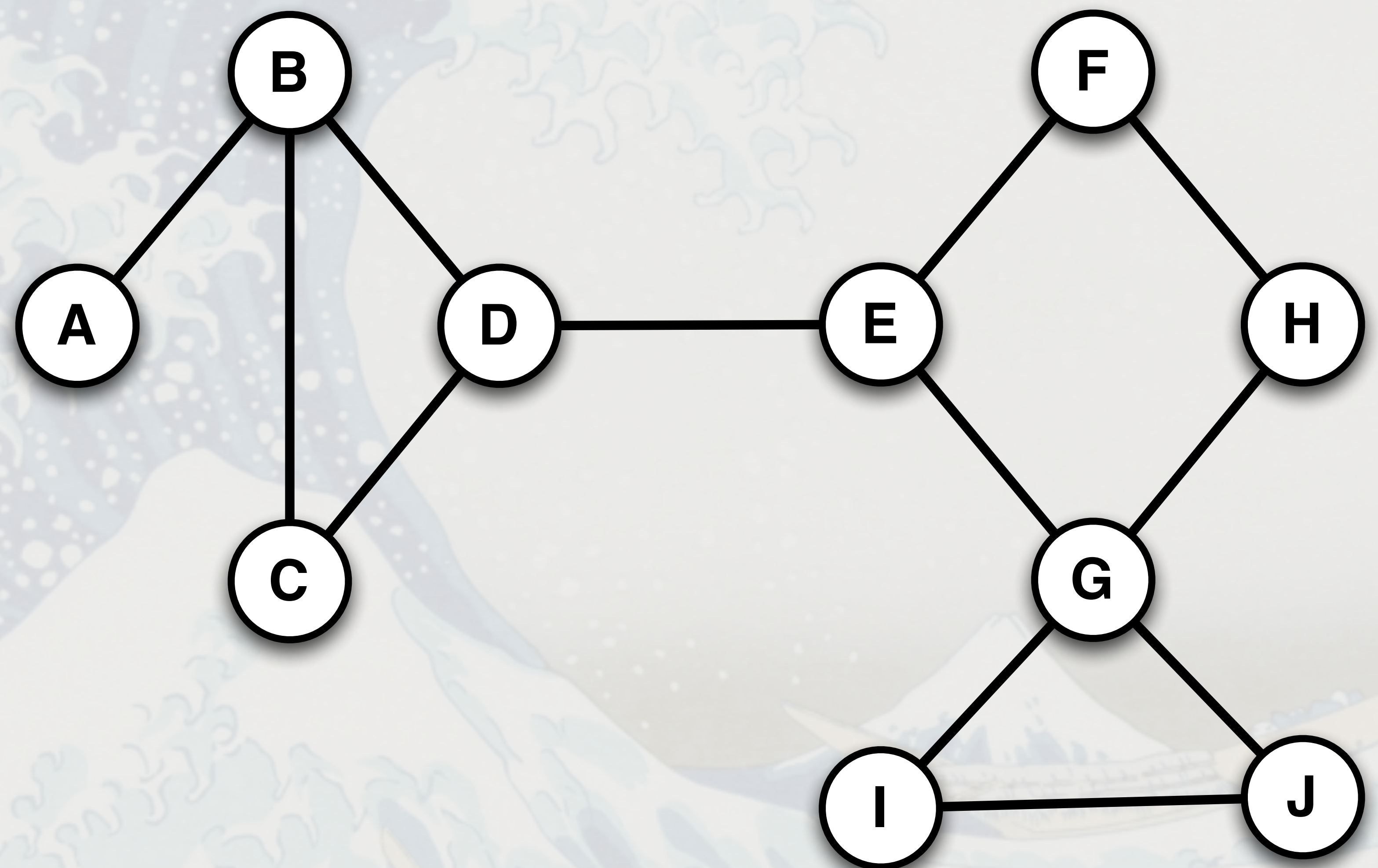
Improvement: from NS to a Minimal Equivalent Graph (MEG)



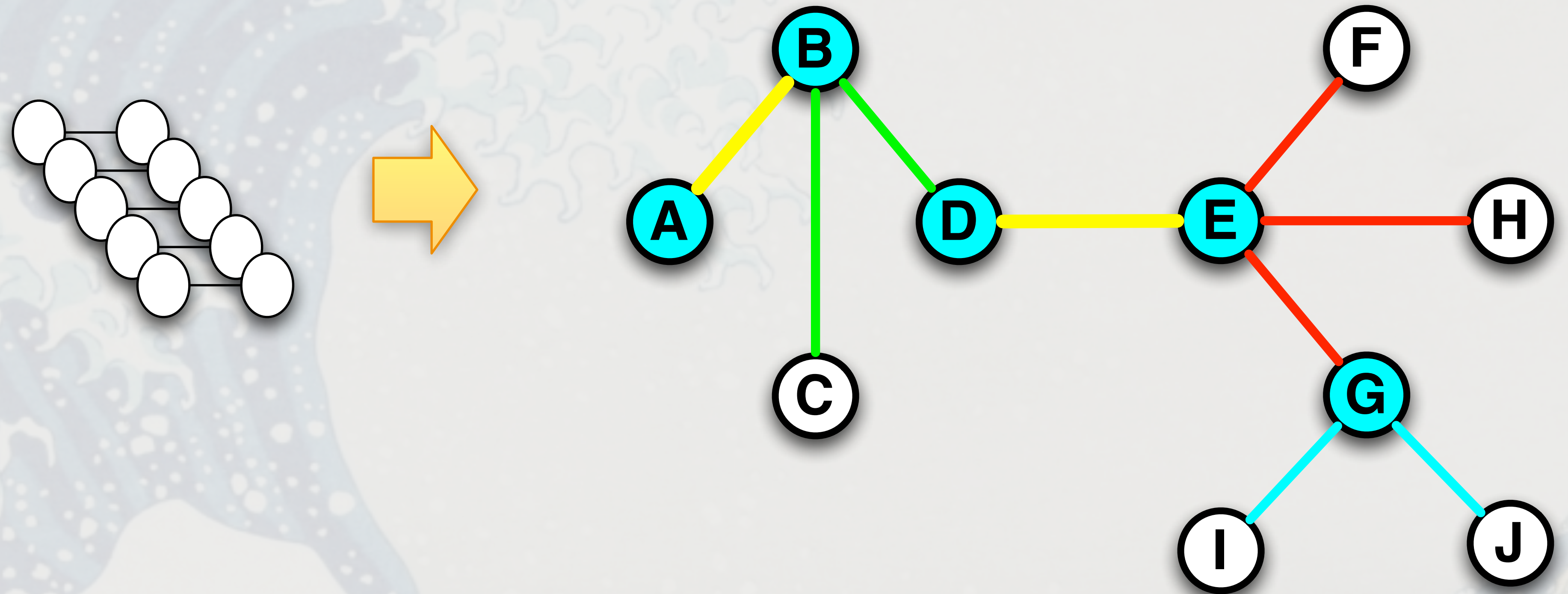
Improvement: from NS to a Minimal Equivalent Graph (MEG)



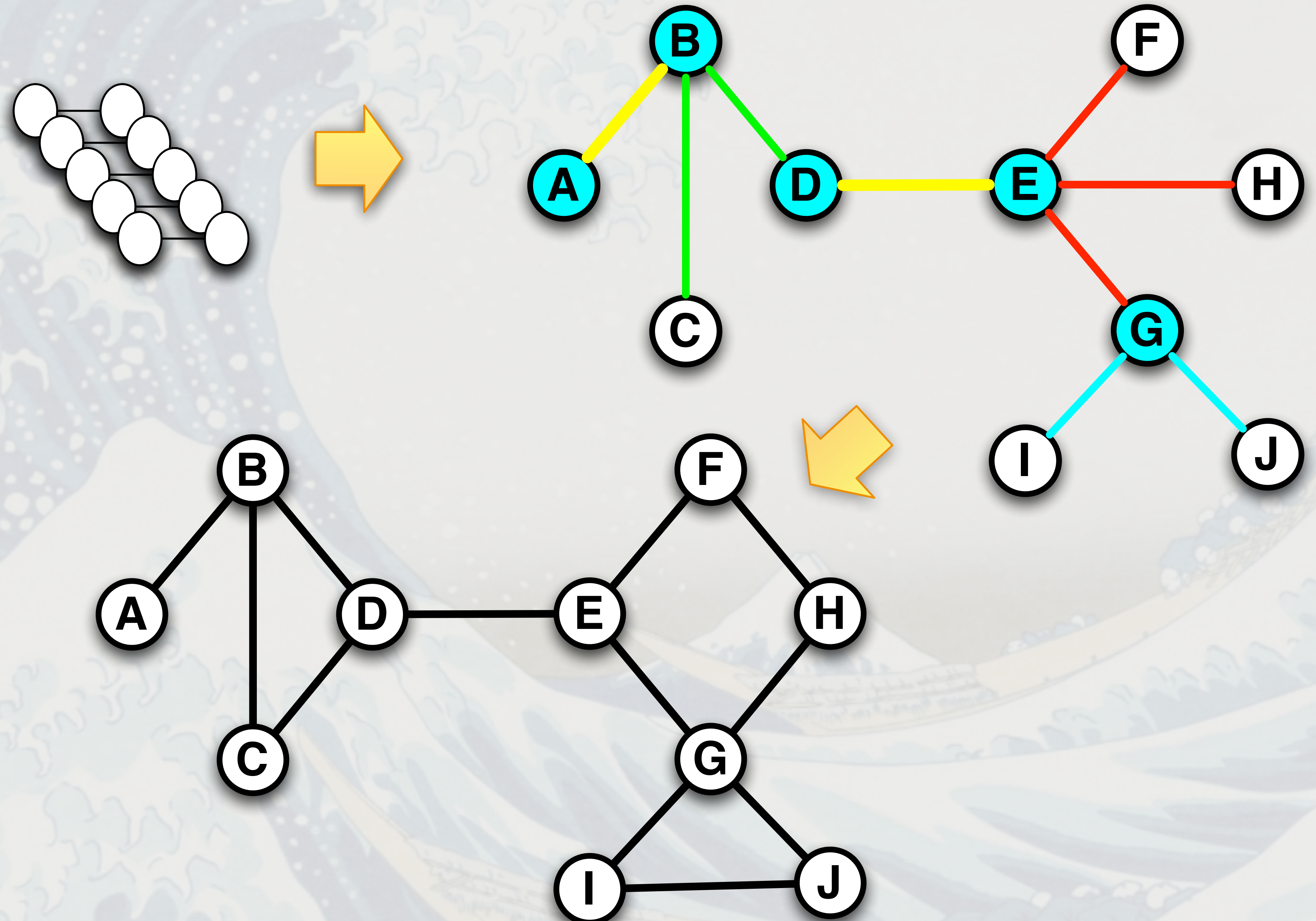
Improvement: from NS to a Minimal Equivalent Graph (MEG)



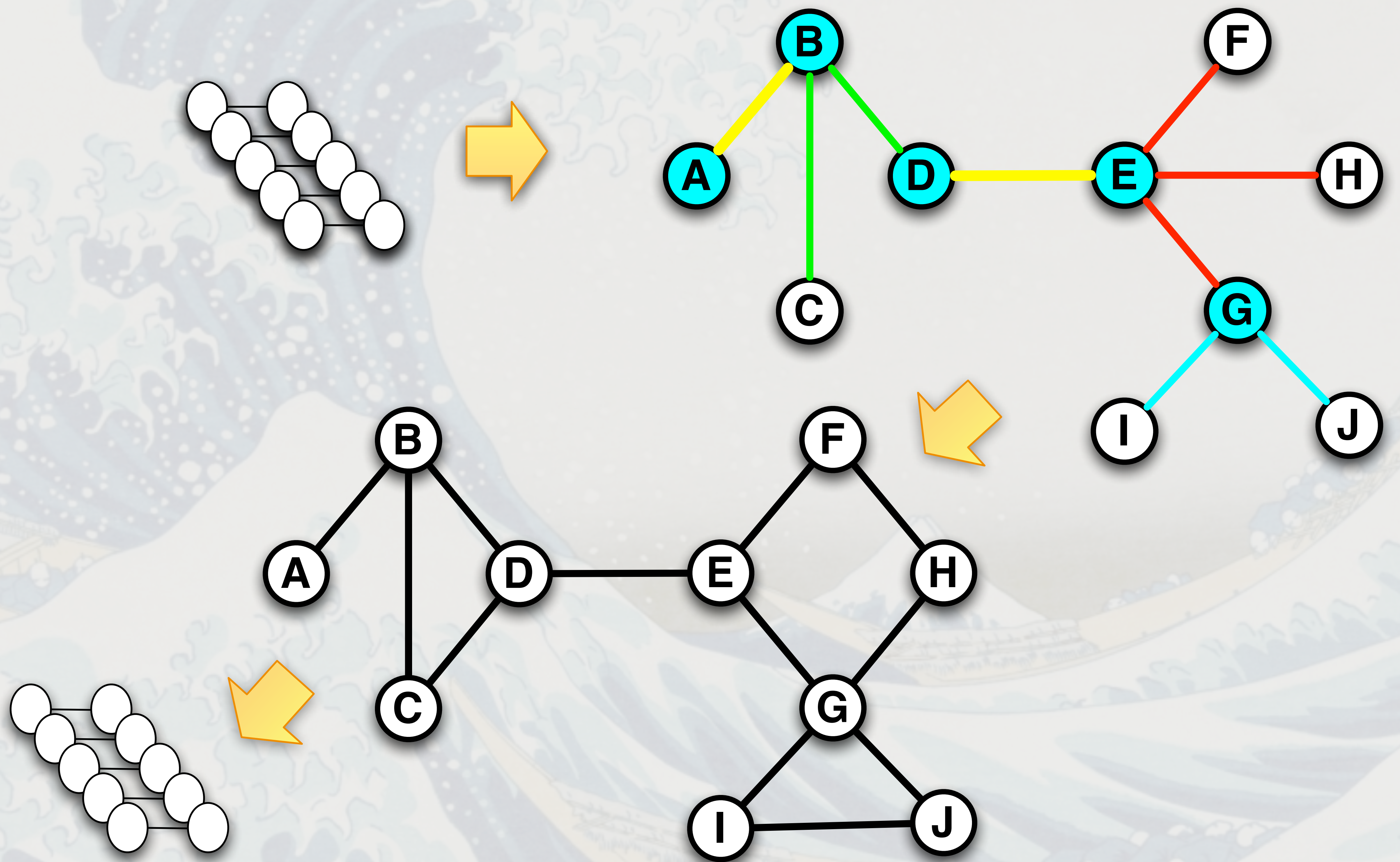
Inside a Reducer



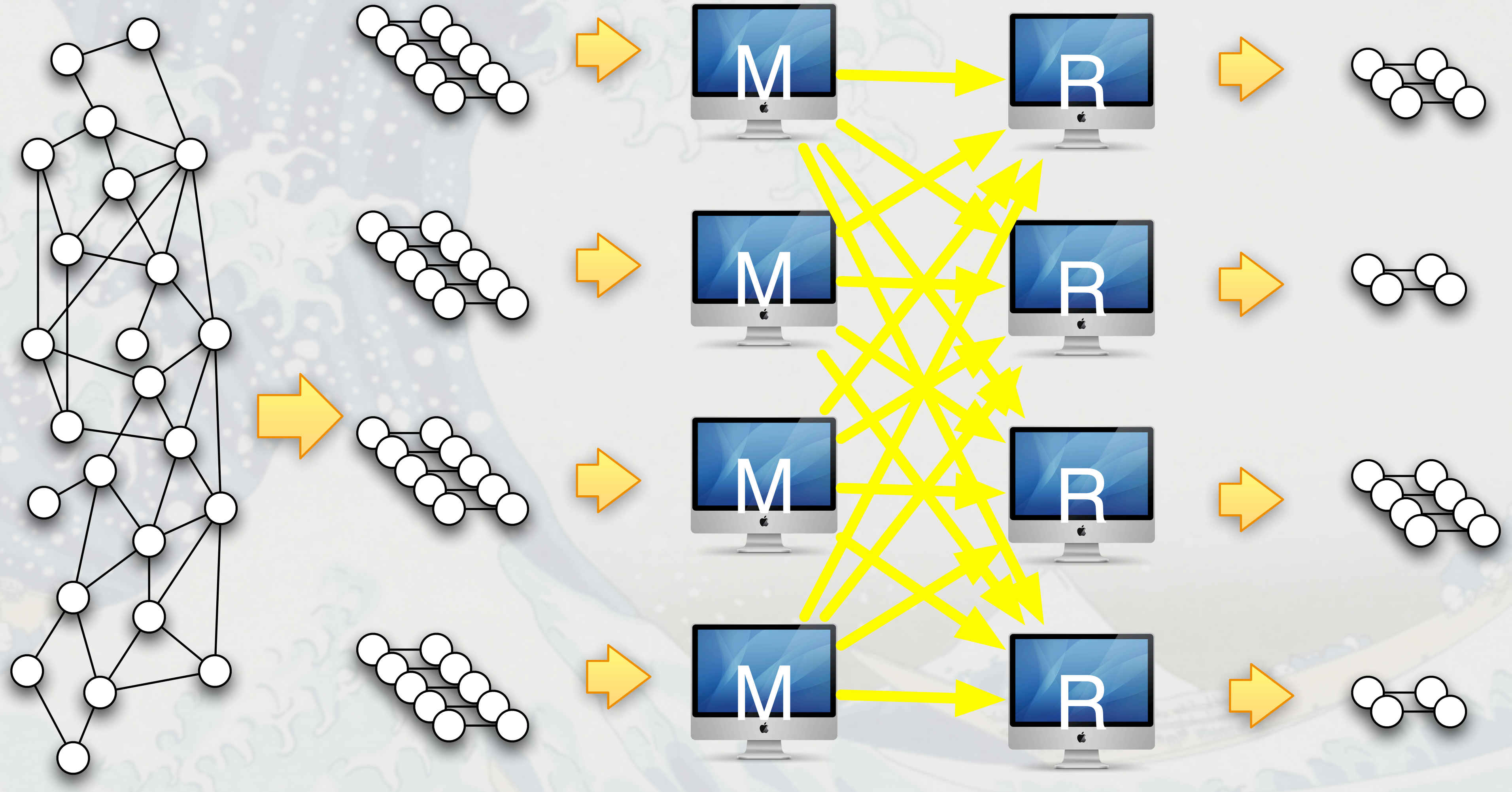
Inside a Reducer



Inside a Reducer

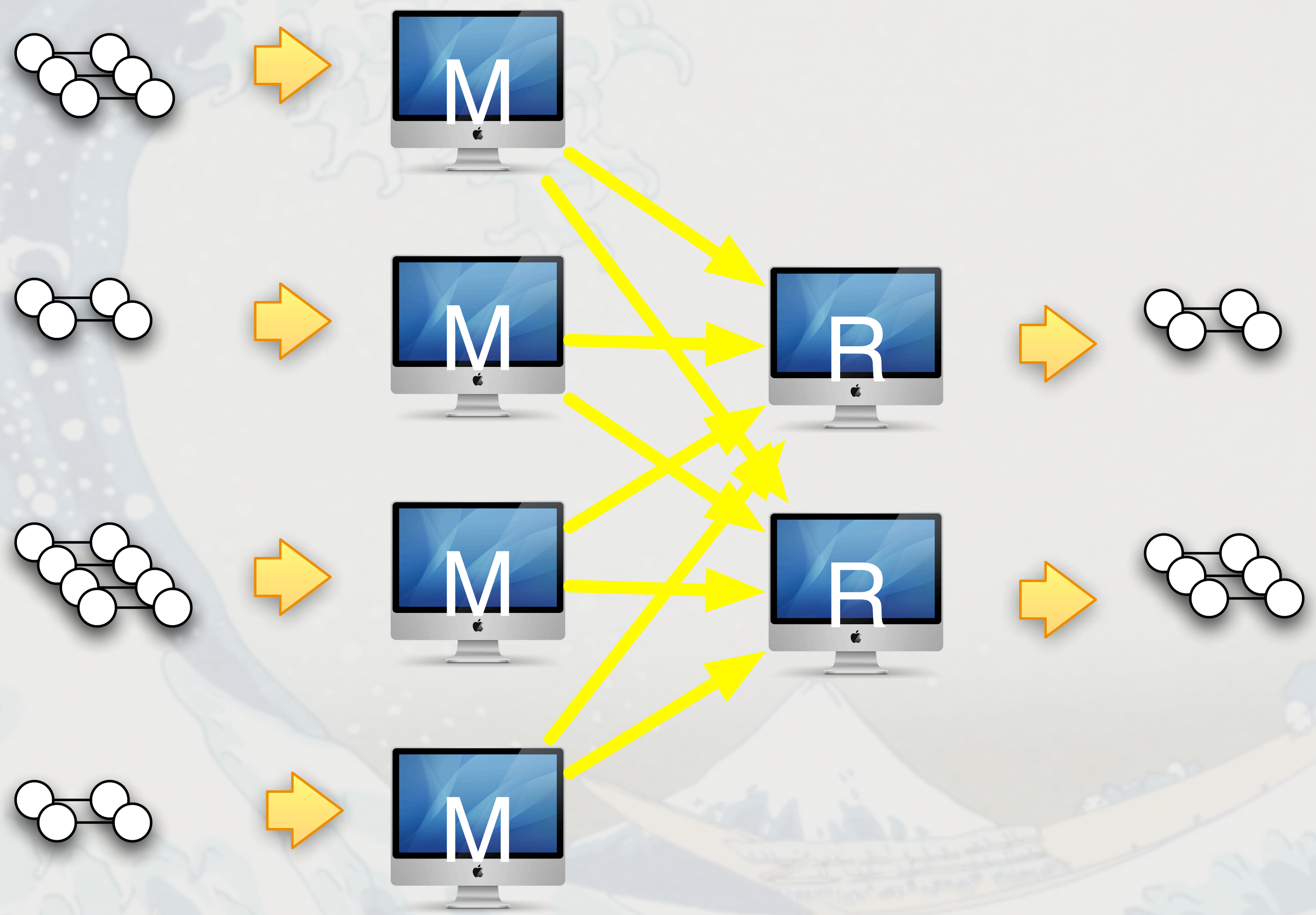


The algorithm



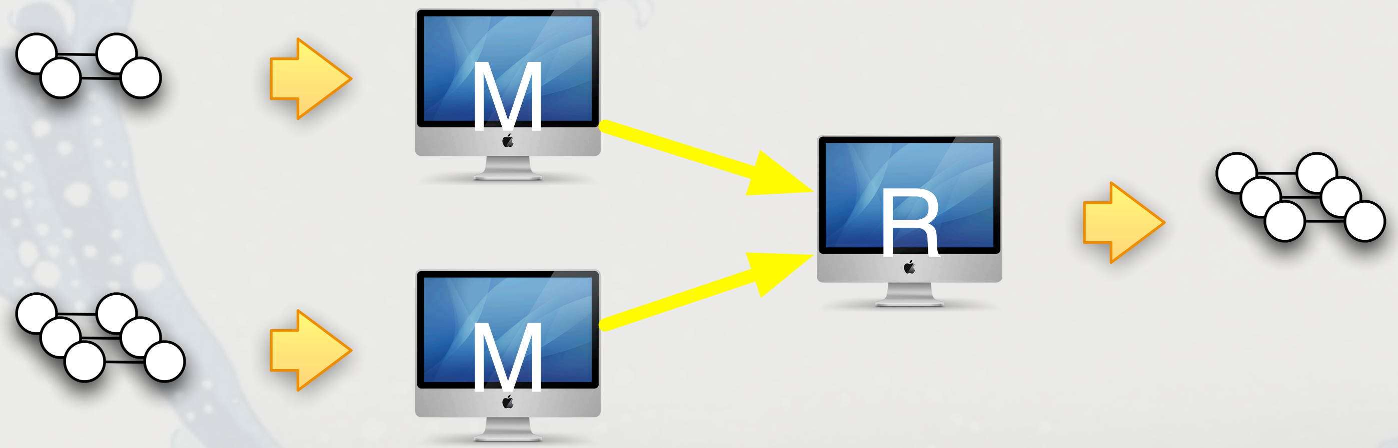
First round

The algorithm



Second round

The algorithm



Third and final round

Remarks...

The use of MEGs allows us to

- have a "consistent" input/output format between mappers and reducers: graph edges (no edge labels/ colors!)
- shuffle the output of each round

Our approach can be seen close to the Filtering technique of Silvio et al. but we do not only filter: we introduce new edges (that replace other edges of the graph).

Current Work

- We are doing some experiments with the MapReduce algorithm seen before (and also with the previous one...)
- We are investigating also the edge partitioning scheme for this algorithm

Thanks!!!

Thanks!!!

Thanks!!!