Supervised Link Prediction in Networks

Alistair Rogers

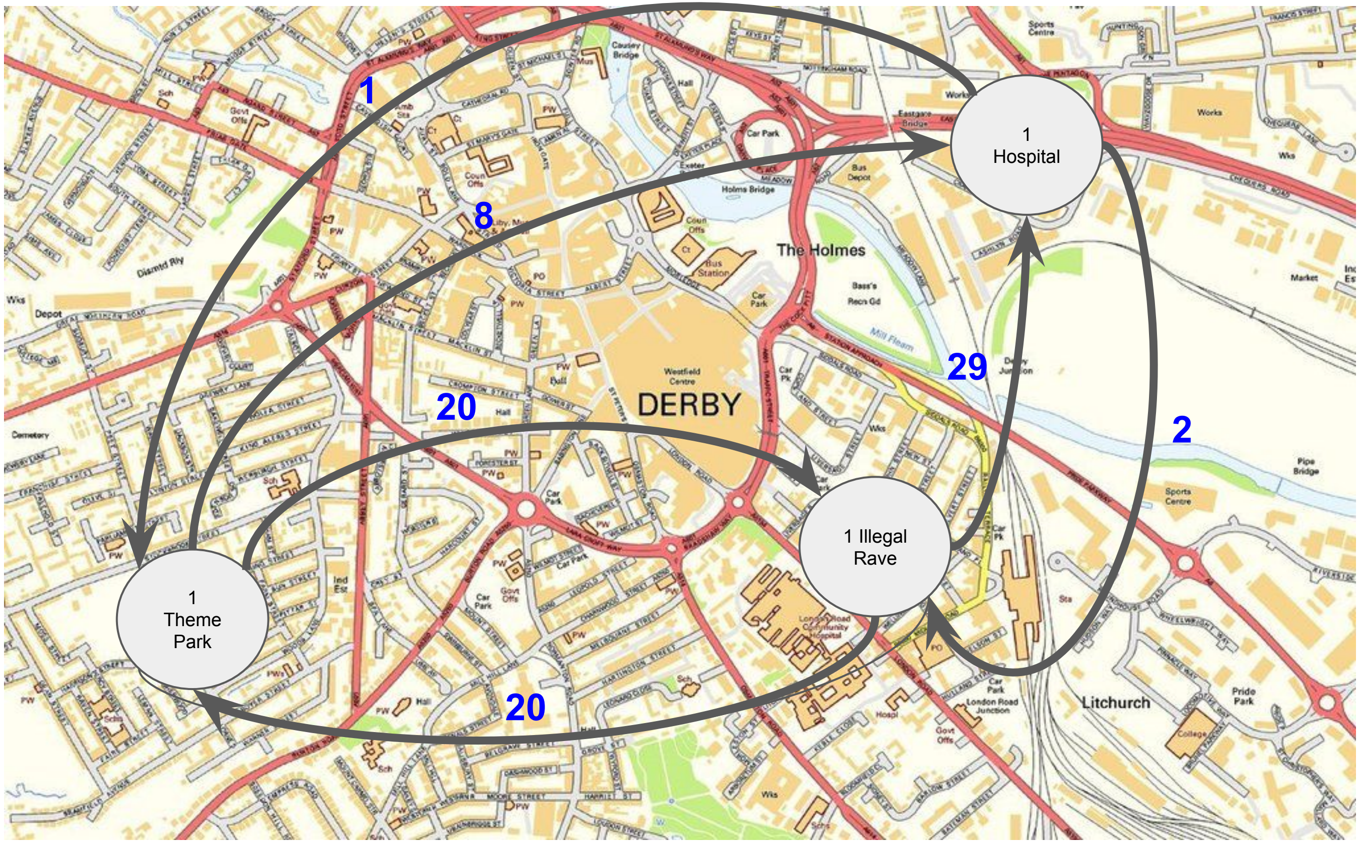
May 2019

About me

- West Swindon born and raised
- Data Scientist - KPMG London (Oct 16 - Jun 18)
- Data Scientist/ML Engineer - OnCorps (Jun 18 - Present)

How did I get involved in this?

Traffic Flow Prediction



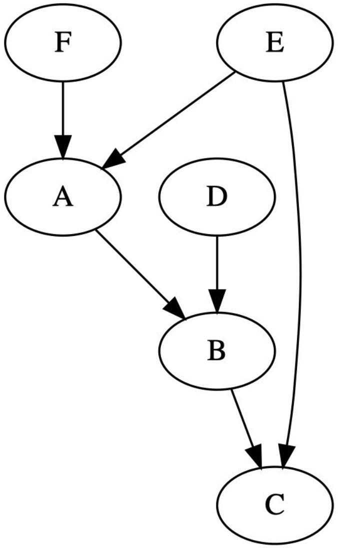
What is a Graph?

From	To
A	B
B	C
D	B
E	C
F	A
E	A

Edgelist

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

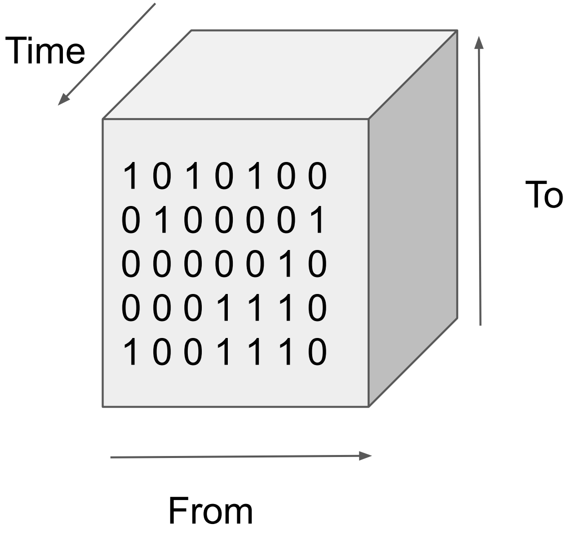
Adjacency Matrix



Visual Representation of Graph

What is a Temporal Graph?

Adjacency Tensor



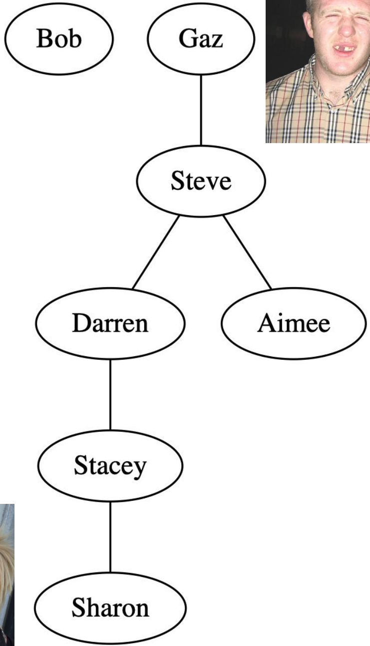
Edgelist with Time Property

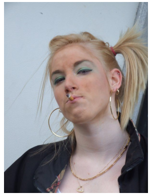
From	To	Time
Frank	Alistair	1
Alistair	Frank	1
Frank	David	1
David	Alistair	2
Frank	David	2

Examples of Temporal Graphs

- Social Networks over time
- Traffic Flows
- IT Failures
- Conflict Networks

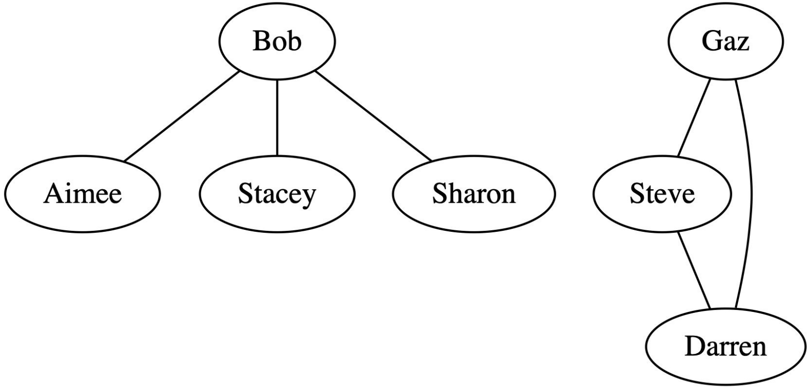
Link Prediction… A Swindon Love Story





$$G_{T_1}$$

$$f(G_{T_1})$$



$$G_{T_2}$$

Sliding Window Approach for Link Prediction

$$\mathbb{P}(E_{T_n}(i,j) | G_{T_{n-1}}, G_{T_{n-2}} \cdots G_{T_{n-k}})$$

Probability of the edge (i, j) at time T_n given the history of the network *G_T_n-1, …, G_T_n-k*

How do you extract features?

Example looking 2 time periods back

Edge	Feature 1 at Time T_n-2	Feature 2 at Time T_n-2	Feature 3 at Time T_n-2	Feature 1 at Time T_n-1	Feature 2 at Time T_n-1	Feature 3 at Time T_n-1	Edge Existence at time T_n
(Bob, Gaz)	0	44	1	55	1	4	1
(Gaz, Steve)	55	0	1.44	4	5	2	1
(Darren, Aimee)	0	0	0	0	11	33	0
(Sharon, Bob)	0	4	1	0	3	1	1

All potential edges

Lagged Features

Target Label

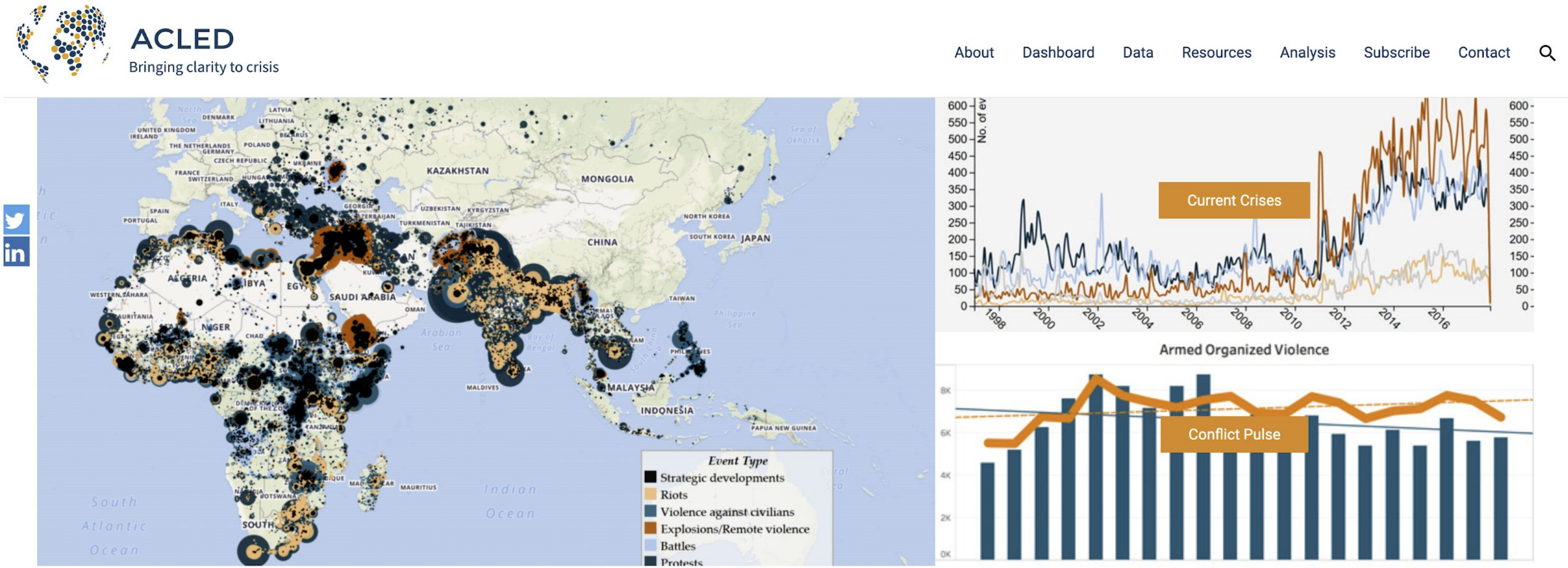
What Features can I extract from a Graph?

- **Jaccard Coefficient of an Edge**
  Proportion of common neighbours relative to total number of neighbours
- **Preferential Attachment of an Edge**
  Probability of a link (i, j) appearing is proportional to the product of the number of neighbours of i and j
- **Resource Allocation of an Edge**
- **Node Based Features**

Logistics + Limitations

- Nodes cannot change over time
- Extremely imbalanced classes (in the classification sense)
- Link prediction metrics in NetworkX (Python) are written in pure python == Slow!!!

ACLED - Can we predict African Conflicts?



ACLED

data_id	iso	event_id_cnty	event_id_no_cnty	event_date	year	time_precision	event_type	sub_event_type	actor1	assoc_actor_1	...	location	latitude	longitude	geo_pre
5022244	180	DRC13977	13977	2019-03-23	2019	1	Riots	Mob violence	Rioters (Democratic Republic of Congo)	Vigilante Group (Democratic Republic of Congo)	...	Bunia	1.5667	30.2500	
5022245	180	DRC13978	13978	2019-03-23	2019	2	Violence against civilians	Attack	Unidentified Armed Group (Democratic Republic ...	NaN	...	Nyiragongo	-1.5219	29.2496	
5022284	404	KEN6858	6858	2019-03-23	2019	1	Protests	Protest with intervention	Protesters (Kenya)	NaN	...	Kibabii	0.6199	34.5275	
5022293	434	LBY7472	7472	2019-03-23	2019	1	Protests	Peaceful protest	Protesters (Libya)	Magarha Ethnic Group (Libya)	...	Tripoli	32.8925	13.1800	
5022311	466	MLI2741	2741	2019-03-23	2019	1	Violence against civilians	Attack	Dan Na Ambassagou	NaN	...	Ogassogou	14.0088	-3.8872	

Conflict Graph

- Nodes defined as group type within country e.g. Ethnic Militia - Angola, Civilians - Ghana
- Edges - all possible conflicts. 1 if at least 1 conflict, 0 if not
- Model predicting edge should score for recall, not accuracy.

    False negatives == very very bad!

Conflict Graph Features + Target (1997-11)

	agent1	agent2	pref_attachment	resource_alloc_com	jaccard_coef
42960	Ethnic militia-Niger	Government or mutinous force-Niger	16	0.00	0.000000
22800	Government or mutinous force-Niger	Government or mutinous force-Niger	16	2.75	1.000000
32811	Ethnic militia-Niger	Ethnic militia-Niger	16	0.75	0.666667
44485	Ethnic militia-Niger	Political militia-South Africa	12	0.00	0.000000
71517	Government or mutinous force-Niger	Political militia-South Africa	12	0.00	0.000000

	agent1	agent2	target	period
0	Ethnic militia-Kenya	Ethnic militia-Kenya	1.0	1997-11
10	Government or mutinous force-Republic of Congo	Government or mutinous force-Republic of Congo	1.0	1997-11
1	Ethnic militia-Niger	Ethnic militia-Niger	1.0	1997-11
16	Political militia-South Africa	Political militia-South Africa	1.0	1997-11
15	Political militia-Somalia	Political militia-Somalia	1.0	1997-11

Conflict Graph - Combined Features + Target

agent1	Ethnic militia–Cameroon
agent2	Protesters–Uganda
pref_attachment_1periods_prev	0
resource_alloc_com_1periods_prev	0
jaccard_coef_1periods_prev	0
pref_attachment_2periods_prev	0
resource_alloc_com_2periods_prev	0
jaccard_coef_2periods_prev	0
pref_attachment_3periods_prev	0
resource_alloc_com_3periods_prev	0
jaccard_coef_3periods_prev	0
pref_attachment_4periods_prev	0
resource_alloc_com_4periods_prev	0
jaccard_coef_4periods_prev	0
pref_attachment_5periods_prev	0
resource_alloc_com_5periods_prev	0
jaccard_coef_5periods_prev	0
pref_attachment_6periods_prev	0
resource_alloc_com_6periods_prev	0
jaccard_coef_6periods_prev	0
pref_attachment_7periods_prev	0
resource_alloc_com_7periods_prev	0
jaccard_coef_7periods_prev	0
pref_attachment_8periods_prev	0
resource_alloc_com_8periods_prev	0
jaccard_coef_8periods_prev	0
pref_attachment_9periods_prev	0
resource_alloc_com_9periods_prev	0
jaccard_coef_9periods_prev	0
pref_attachment_10periods_prev	0
resource_alloc_com_10periods_prev	0
jaccard_coef_10periods_prev	0
pref_attachment_11periods_prev	0
resource_alloc_com_11periods_prev	0
jaccard_coef_11periods_prev	0
pref_attachment_12periods_prev	0
resource_alloc_com_12periods_prev	0
jaccard_coef_12periods_prev	0
target	0
period	1998–1

Results

Model: Balanced Random Forest

```
[[17777989   698892]
 [     950     8889]]
Recall Score (TRAIN): 0.9034454721008233
F1 (TRAIN): 0.024773557035757086
```

```
[[829635  93652]
 [    90    959]]
Recall Score (TEST): 0.9142040038131554
F1 (TEST): 0.020050177712732594
```

```
Feature ranking:
1. Feature: pref_attachment_1periods_prev 0.14901047971698075
2. Feature: pref_attachment_2periods_prev 0.10556877419792446
3. Feature: pref_attachment_5periods_prev 0.0948233891897029
4. Feature: pref_attachment_3periods_prev 0.08363538078858798
5. Feature: pref_attachment_7periods_prev 0.06590819953750122
6. Feature: pref_attachment_6periods_prev 0.05290802920080257
7. Feature: pref_attachment_4periods_prev 0.04488638263978326
8. Feature: pref_attachment_8periods_prev 0.0388324366070451
9. Feature: pref_attachment_12periods_prev 0.0370735838149487
10. Feature: pref_attachment_9periods_prev 0.0345007636167276
```