

Roberto Franzosi

Emory University

Quantifying Narrative

An Application to the Rise of Italian Fascism (1919-1922)

Information Beyond Shannon

Venice, Istituto Veneto, December 29-30, 2008

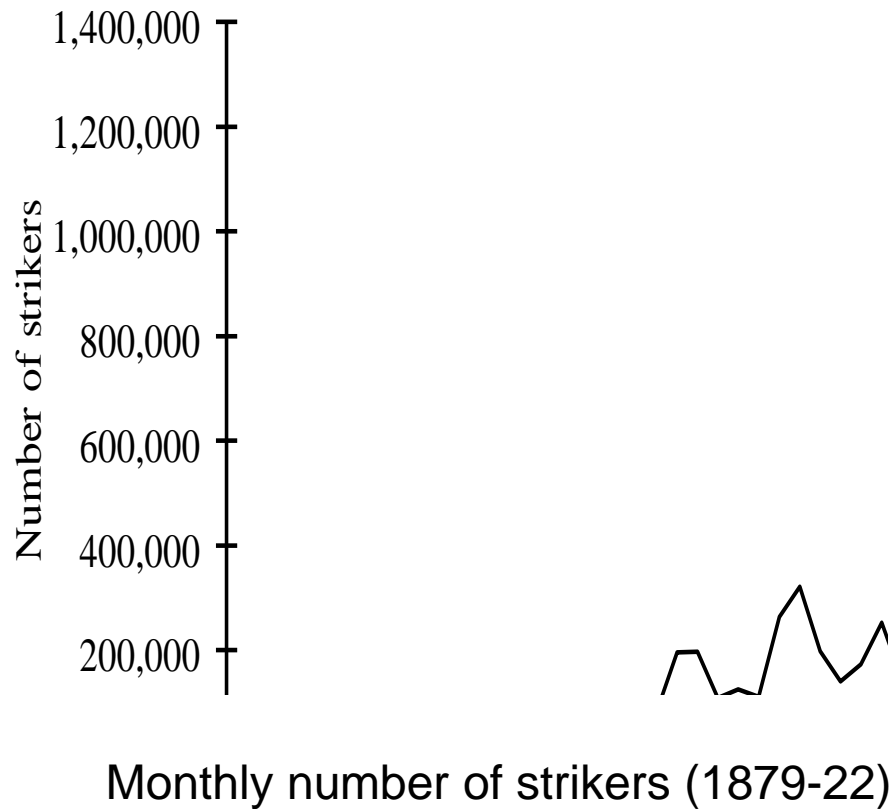
The social science perspective

Information *BEFORE* Shannon

From words to numbers

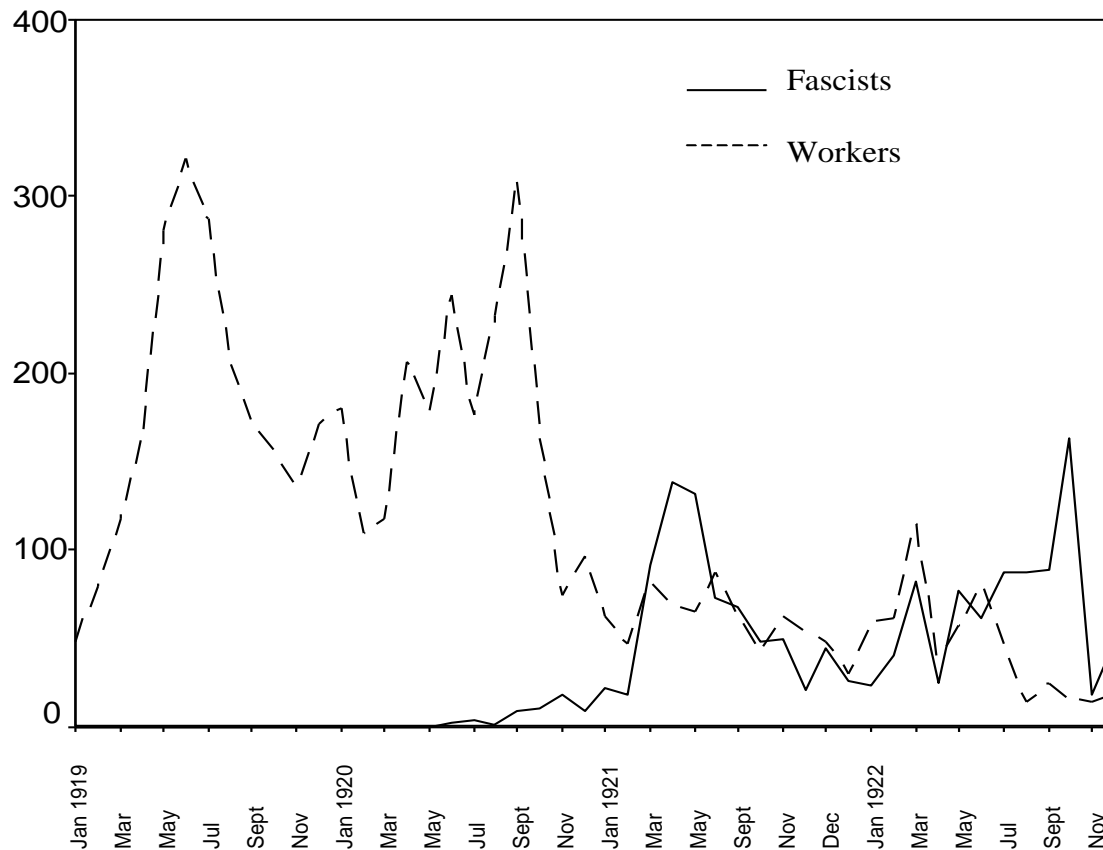
In the 1980s I got involved in computer-assisted parsing of narrative. Storing elements of narrative (who, what, when, where, why, or “story grammar”) in a RDBMS allowed me to go “from words to numbers.” I applied the technique to some 60,000 newspaper articles on the rise of Italian fascism (1919-1922).

Why the 1919-22 period?



“Red Years” and “Black Years” (1919-22)

Working-class mobilization and fascist counter-mobilization

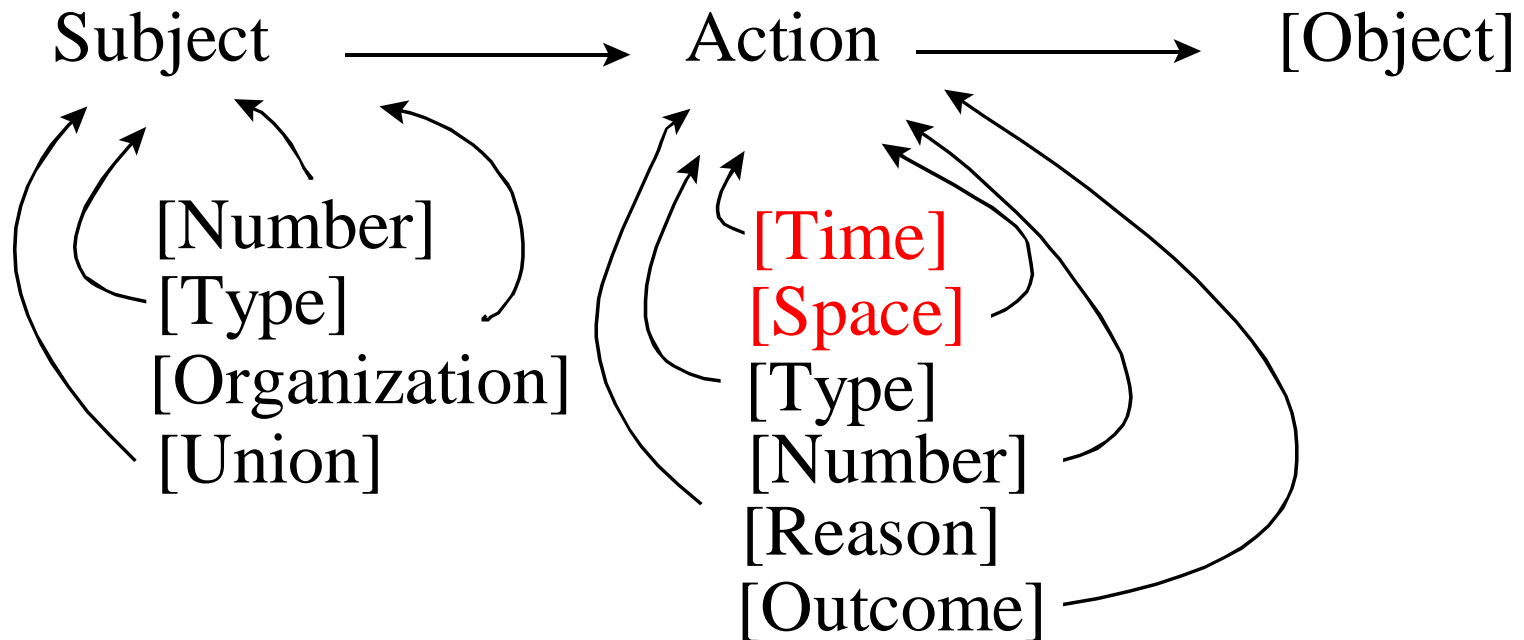


Typical narrative text

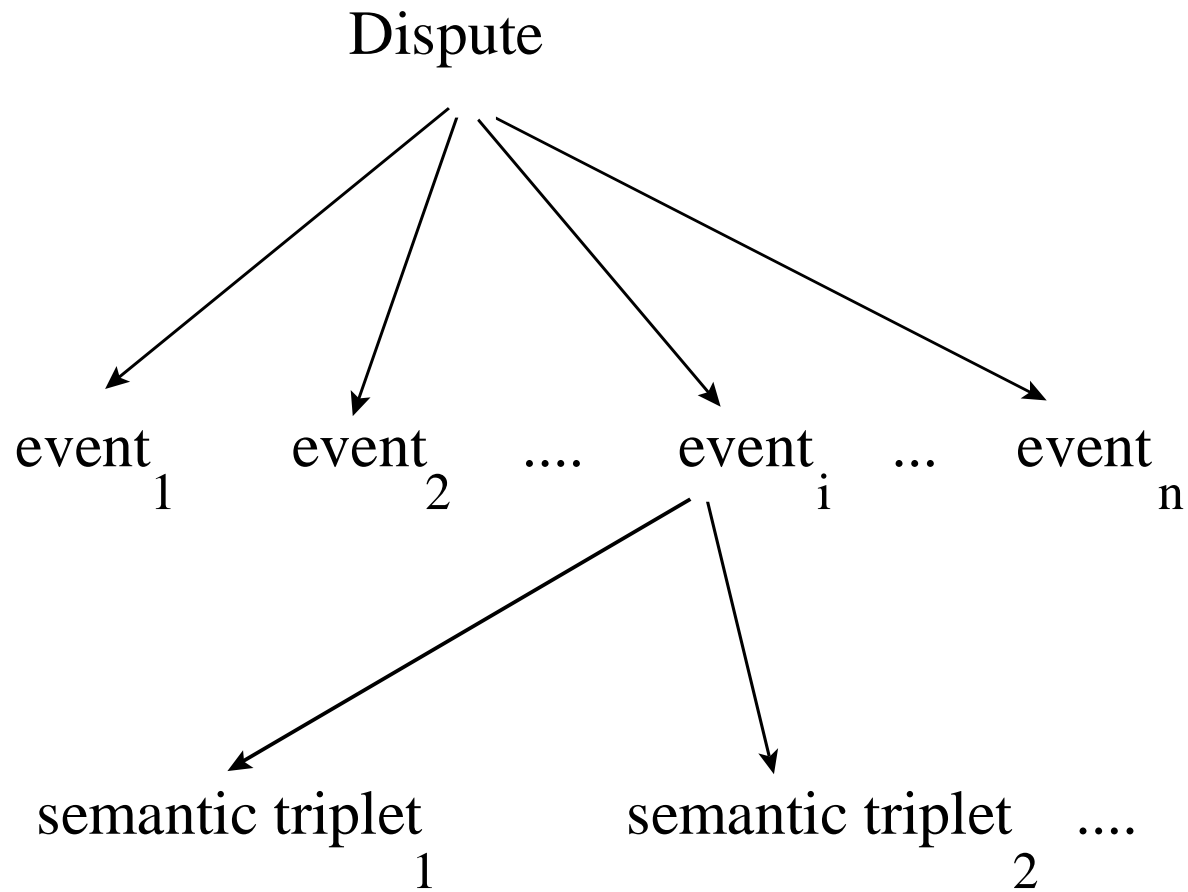
Fascists arrive on a truck on the night of 5/7/1921 in Bissone di S. Cristina at the pub of Mr. Prati. There, they seize 5 workers and take them outside. They beat up the 5 workers with retorted cowhide with lead inside.

Story grammar

(“semantic triplet”)



Hierarchical structure



Rewrite rules

<dispute>	→	{<event>}
<event>	→	{<semantic triplet>}
<triplet>	→	{<subject> {<action>} [{<object>}]}
<subject>	→	{<actor>} [{<modifier>}]
<actor>	→	{<individual>} {<set of individuals>} {<institution>}
...		
<modifier>	→	{[<type>] [<organization>] [<number>] ...}
<type>	→	(male young immigrant skilled ...)
...		

Rewrite rules

1: Legenda:

2:

3: → Rewrite/Production Rule: the object on the left is “rewritten” in terms of the object(s) to the right of the arrow

4: < > demarcates the boundary of an object's name

5: [] indicates that an object is optional

6: { } indicates that an object may have multiple instances

7: ++ denotes One-To-Many (Hierarchical) complex objects

8: + denotes One-To-Few complex objects, i.e. objects made up of a combination of simplex and complex objects

9:

10:

11: Articolo di giornale → <Nome del giornale>

12: <Data del giornale>

13: [<Titolo della pagina>]

14: <Numero pagina>

15: <Numero colonna>

16: <Posizione nella colonna>

17: [<Città dell'articolo: Nome Città>]

18: [<Data dell'articolo: Data del giornale>]

19: <Potenzialmente qualifica>

20: <Verificato IO>

21: [<Verificato IO Xref>]

22:

23:

24: <+Vertenza> → <Tipo di evento> {[+<Spazio>]} {[<Sindacato>]} {[<Mestiere>]} {[<Azienda>]} {[<Settore>]}

25: {[<Ragione: Nome ragione/rivendicazione>]} [<Verificato CS>] [<Nuove Xrefs>]

26: {[<NON USARE! (Città): Nome Città>]} {[++<Sotto Vertenza>]}

27: <Tipo di evento> → accordo | agitazioni | corteo | sciopero | violenza | ...

28:

29: <+Spazio> → {[+<Città>]} {[+<Territorio>]}

30:

31: <+Città> → [<Direzione spaziale>] [<Nome città>] [<Provincia>] [+<Località nella città>] [+<Aggiungi città>]

32: <Direzione spaziale> → a | attraverso | da | per | verso | ...

33: <Nome città> → Acqui | Ancona | Benevento | Borgo S. Donnino | Lucera | ...

34: <Provincia> → alessandria | Asti | bologna | catania | Firenze | ...

35:

36: <+Località nella città> → [<Qualificatore della località>] [<Frazione>] [<Quartiere>] [<Strada>] [<Piazza>]

37: [<Altro riferimento cittadino>] [<Numero civico>] [+<Sede di>]

38: [<NON USARE (Ex Sede di): Sede di>] [+<Aggiungi località>]

39: <Qualificatore della località> → davanti | fuori da | lontano da | presso | vicino a | ...

40: <Frazione> → fontana predosa | lendosio | sette camini | ...

41: <Quartiere> → bicocca | Capanne | gabbiana | S. Giacomo | Santa Maria Novella | ...
42: <Strada> → Corso Cavour | corso siccardi | Via Campo Lodigiano | via Carmine | via zola | ...
43: <Piazza> → Canossa | cavour | Mazzini | S.Francesco | san marco | ...
44: <Altro riferimento cittadino> --> centro città | abitazione | abitazione di barrotti roberto |
45: abitazione di castrucci augusto | abitazione di cocci | ...
46: <Numero civico> → 3 | 7 | 10 | 18 | 22 | ...
47:
48: <+Sede di> → [+<Organizzazione>]
49:
50: <+Organizzazione> → [+<Nome di istituzione>] [<NON USARE!!!!: Azienda>] {[+<Organizzazione complessa>]}
51:
52: <+Nome di istituzione> → [<Azienda>] [<Sindacato>] [<Partito politico>] [<Organismo dello stato>]
53: [<Altre organizzazioni>]
54: <Azienda> → romeo | Agnesi | Ansaldo S. Giorgio | Banfi | beretta | ...
55: <Sindacato> → camera del lavoro | Associazione industriali del legno |
56: federazione degli agricoltori | federazione dei chimici | FIOM -CAMERA DEL LAVORO | ...
57: <Partito politico> → arditi del popolo | fascio di combattimento | partito fascista |
58: partito popolare | partito socialista | ...
59: <Organismo dello stato> → Arma dei carabinieri | camera dei deputati | consiglio dei ministri |
60: Corte d'assise | ministero degli interni | ...
61: <Altre organizzazioni> → Associazione Mutilati di Guerra | Cooperativa del Vino | agraria |
62: alloggi | amici dell'arte | ...
63: <NON USARE!!!!: Azienda> → romeo | Agnesi | Ansaldo S. Giorgio | Banfi | beretta | ...
64:
65: <+Organizzazione complessa> → [+<Numero e livello dell'unità organizzativa>] [+<Nome dell'unità>]
66: [<Organo direttivo>] [+<Numero di individui nell'unità>]
67: {[+<Localizzazione dell'unità>]} {[+<Proprietà>]} [+<Aggiungi unità>]
68:
69: <+Numero e livello dell'unità organizzativa> → [+<Numero di unità: Numero>] <Livello dell'unità>
70:
71: <+Numero> → [<Qualificatore approssimativo>] [<Qualificatore comparativo>]
72: [+<Valore qualitativo>] [+<Valore quantitativo>]
73: <Qualificatore approssimativo> → circa | forse | piu' o meno | probabilmente | quasi | ...
74: <Qualificatore comparativo> → almeno | fino a | meno | più di | poco più di | ...
75:
76: <+Valore qualitativo> → [<Qualificatore di quantità>] [<Numerale>]
77: <Qualificatore di quantità> → alcuni | molti | numerosi | parecchi | pochi | ...
78: <Numerale> → centinaio | decina | dozzina | migliaia | trentina | ...
79:
80: <+Valore quantitativo> → [<Valore singolo: Valore numerico>] [+<Intervallo di valori>]
81: [+<Valore sul totale>]

82: <Valore singolo: Valore numerico> → 1 | 2 | 10 | 14 | 25 | ...

83:

84: <+Intervallo di valori> → <Valore inferiore: Valore numerico>

85: <Valore superiore: Valore numerico>

86: <Valore inferiore: Valore numerico> → 1 | 2 | 10 | 14 | 25 | ...

87: <Valore superiore: Valore numerico> → 1 | 2 | 10 | 14 | 25 | ...

88:

89: <+Valore sul totale> → <Valore numerico> <Sul totale: Valore numerico>

90: <Valore numerico> → 1 | 2 | 10 | 14 | 25 | ...

91: <Sul totale: Valore numerico> → 1 | 2 | 10 | 14 | 25 | ...

92: <Livello dell'unità> → divisione | fabbrica | reparto | settore | ufficio | ...

93:

94: <+Nome dell'unità> → [+<Nome di istituzione>] [<Reparto>] [<Ufficio>] [<Fabbrica>] [<Divisione>]

95: [<Gruppo/Holding>] { [<Settore>] }

96: <+Nome di istituzione> → Re-write rules for this object on line 52

97: <Reparto> --> armi leggere | assemblaggio | Brigata Sassari | motori | verniciatura | ...

98: <Ufficio> → amministrativo | del personale | finanze | pianificazione | tecnico | ...

99: <Fabbrica> → fiammiferi | stabilimenti | ...

100: <Divisione> → 38° fanteria | 8° fanteria | fanteria | Livorno | pubblica sicurezza | ...

101: <Gruppo/Holding> → String

102: <Settore> → agrario | alberghiero | calzaturiero | estrattivo | metallurgico | ...

103: <Organo direttivo> → amministrazione | Comitato centrale | Comitato centrale di agitazione |

104: comitato d'azione | commissione di rappresentanti | ...

105:

106: <+Numero di individui nell'unità> → { [+<Numero>] }

107: <+Numero> → Re-write rules for this object on line 71

108:

109: <+Localizzazione dell'unità> → { [+<Spazio>] }

110: <+Spazio> → Re-write rules for this object on line 29

111:

112: <+Proprietà> → { [+<Attore>] } { [+<Organizzazione>] }

113:

114: <+Attore> → { [+<Individuo>] } { [+<Attore collettivo>] } { [+<Organizzazione>] }

115:

116: <+Individuo> → <Nome di attore individuale> { [+<Caratteristiche personali>] }

117: <Nome di attore individuale> → anarchico | carabiniere | cittadino | comunista |

118: Individuo | ...

119:

120: <+Caratteristiche personali> → [+<Nome e cognome>] [+<Relazione di parentela>]

121: { [+<Genere>] } { [+<Età>] } { [+<Residenza>] } [<Nazionalità>]

122: {{<Tipo di attore>}} {{<Mestiere>}} {{<Settore>}}
 123: {{<Qualifica>}} {{+<Organizzazione>}}
 124: [<Ruolo nell'organizzazione>]
 125: {{<Affiliazione sindacale: Sindacato>}}
 126: {{<Affiliazione politica: Partito politico>}}
 127: {{<Parte del corpo>}} [+<Sottinsieme (di)>]
 128:
 129: <+Nome e cognome> → [<Nome>] [<Cognome>]
 130: <Nome> → Adriano | Gennaro | raffaele | umberto | Vittorio | ...
 131: <Cognome> → Abate | Barbato | Battaglia | Ceresa | Gamberini | ...
 132:
 133: <+Relazione di parentela> → <Parentela> +<Attore>
 134: <Parentela> → amico | cognato | figlio | fratello | madre | ...
 135: <+Attore> → Re-write rules for this object on line 114
 136: <Genere> → Femmina | Maschio | ...
 137: <Età> → 24 | 42 | 60 | anziano | giovane | ...
 138:
 139: <+Residenza> → {{+<Spazio>}}
 140: <+Spazio> → Re-write rules for this object on line 29
 141: <Nazionalità> → austriaca | francese | inglese | serba | tedesca | ...
 142: <Tipo di attore> → accusati | affittuari | crumiro | disoccupati | minorenni | ...
 143: <Mestiere> → lavoratori parrucchieri | ferrovieri | fiammiferai | fornai | operai | ...
 144: <Settore> → agrario | alberghiero | calzaturiero | estrattivo | metallurgico | ...
 145: <Qualifica> → capi officine | capi tecnici | capo servizio | terza categoria |
 146: arciprete | ...
 147: <+Organizzazione> → Re-write rules for this object on line 50
 148: <Ruolo nell'organizzazione> → amministratore | capo ufficio | consigliere delegato |
 149: direttore | segretario nazionale | ...
 150: <Affiliazione sindacale: Sindacato> → camera del lavoro |
 151: Associazione industriali del legno |
 152: federazione degli agricoltori |
 153: federazione dei chimici |
 154: FIOM -CAMERA DEL LAVORO | ...
 155: <Affiliazione politica: Partito politico> → arditi del popolo |
 156: fascio di combattimento |
 157: partito fascista | partito popolare |
 158: partito socialista | ...
 159: <Parte del corpo> → braccio | cuore | gamba | mano | orecchio | ...
 160:
 161: <+Sottinsieme (di)> → [+<Attore>]
 162: <+Attore> → Re-write rules for this object on line 114
 163:
 164: <+Attore collettivo> → <Nome di attore collettivo> {{+<Caratteristiche collettive>}}
 165: <Nome di attore collettivo> → anarchici | assessori | autorità | disoccupati |

PC-ACE

Program For Computer-Assisted Coding of Events
(available for free download at www.pc-ace.com)



My data: PC-ACE coded output

Semantic triplet: [Subject: [Actor... (Name of collective actor: **FASCISTS**)]
[Action: (Verb phrase: **ARRIVE**) [Time... (Definite date: **05/07/1921**)]
[Space..City: (City name: **BISSONE DI S. CRISTINA**) (Locality: [Physical
object: (Name of physical object: **PUB**) [Ownership...Individual: (Name of
individual actor: **INDIVIDUAL**) [Personal characteristics... (Last name:
PRATI)]]]]

...

Semantic triplet: [Subject: [Actor... (Name of collective actor: **FASCISTS**)]
[Action: (Verb phrase: **SEIZE**) [Object: [Actor..Collective actor: (Name of
collective actor: **WORKERS**) [Collettive characteristics.... (Single value: **5**)]]]]

Semantic triplet: [Subject: [Actor... (Name of collective actor: **FASCISTS**)]
[Action: (Verb phrase: **TAKE OUTSIDE**) [Action: (Verb phrase: **BEAT UP**)
[Instrument.. (Type of instrument: **RETORTED COWHIDE WITH LEAD**)]
[Object: [Actor..Collective actor: (Name of collective actor: **WORKERS**)
[Collettive characteristics.... (Single value: **5**)]]]]

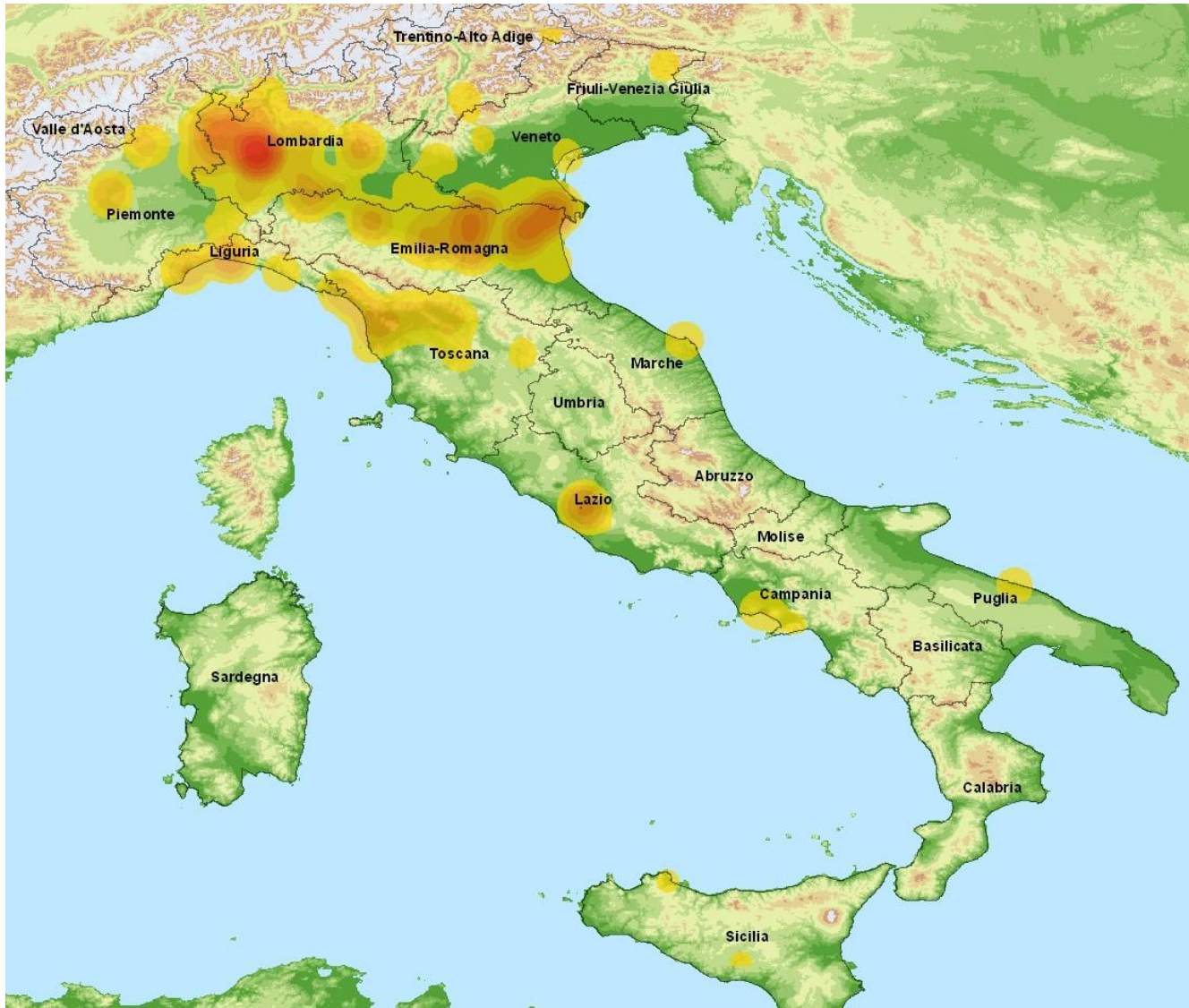
Sources and data

Newspaper name	<i>Il Lavoro</i>	<i>Avanti!</i>	<i>Popolo d'Italia</i>
No. of articles	15,400	17,894	19,871
No. of triplets	17,601	139,530	93,558
Year completion of data collection	1990	2006	2008

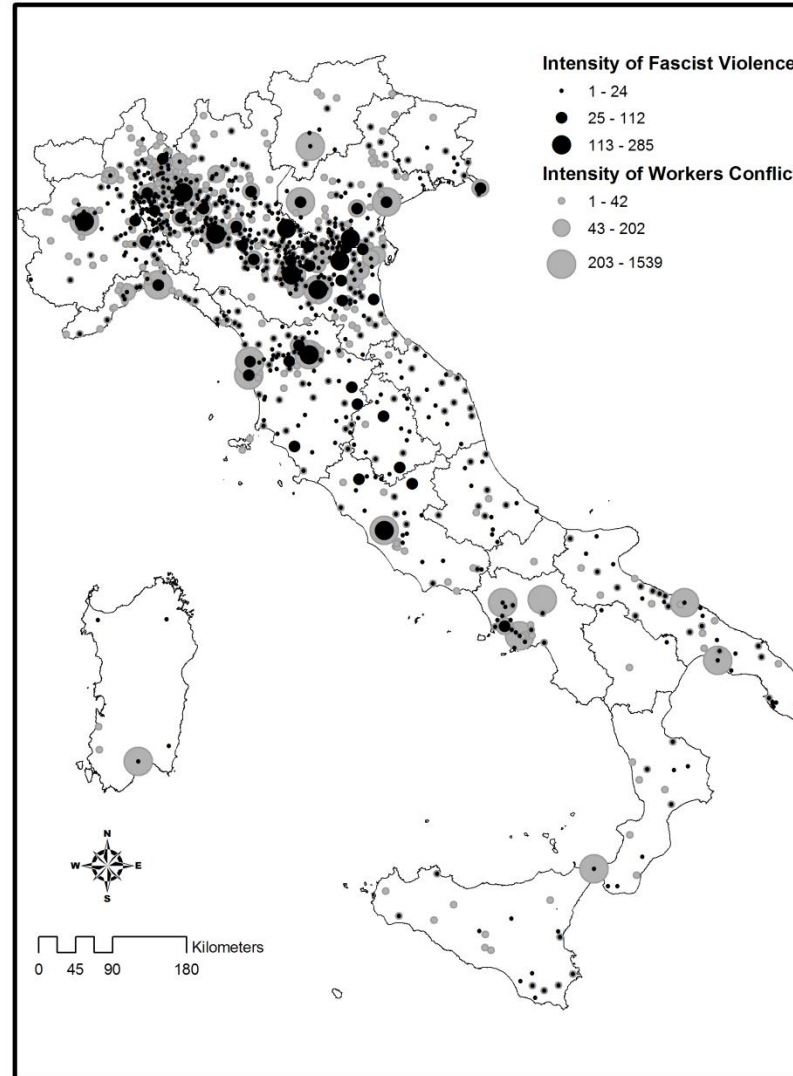
Acknowledgments

1. University of Wisconsin-Madison
2. National Science Foundation
3. University of Reading
4. University of Trento
5. Provincia Autonoma di Trento
6. Emory University

Geography of *Conflict & Violence*

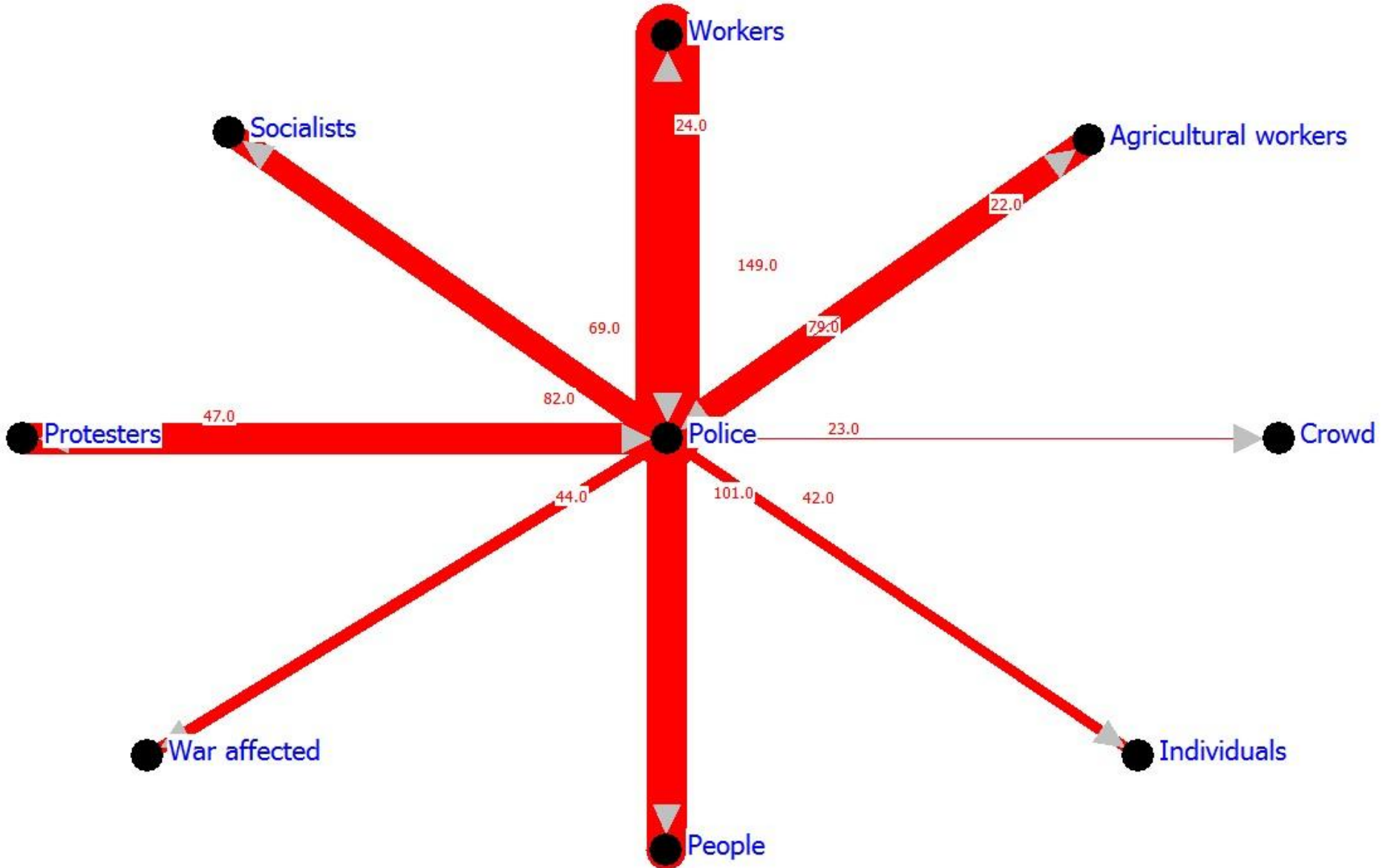


Geography of *Conflict & Violence*



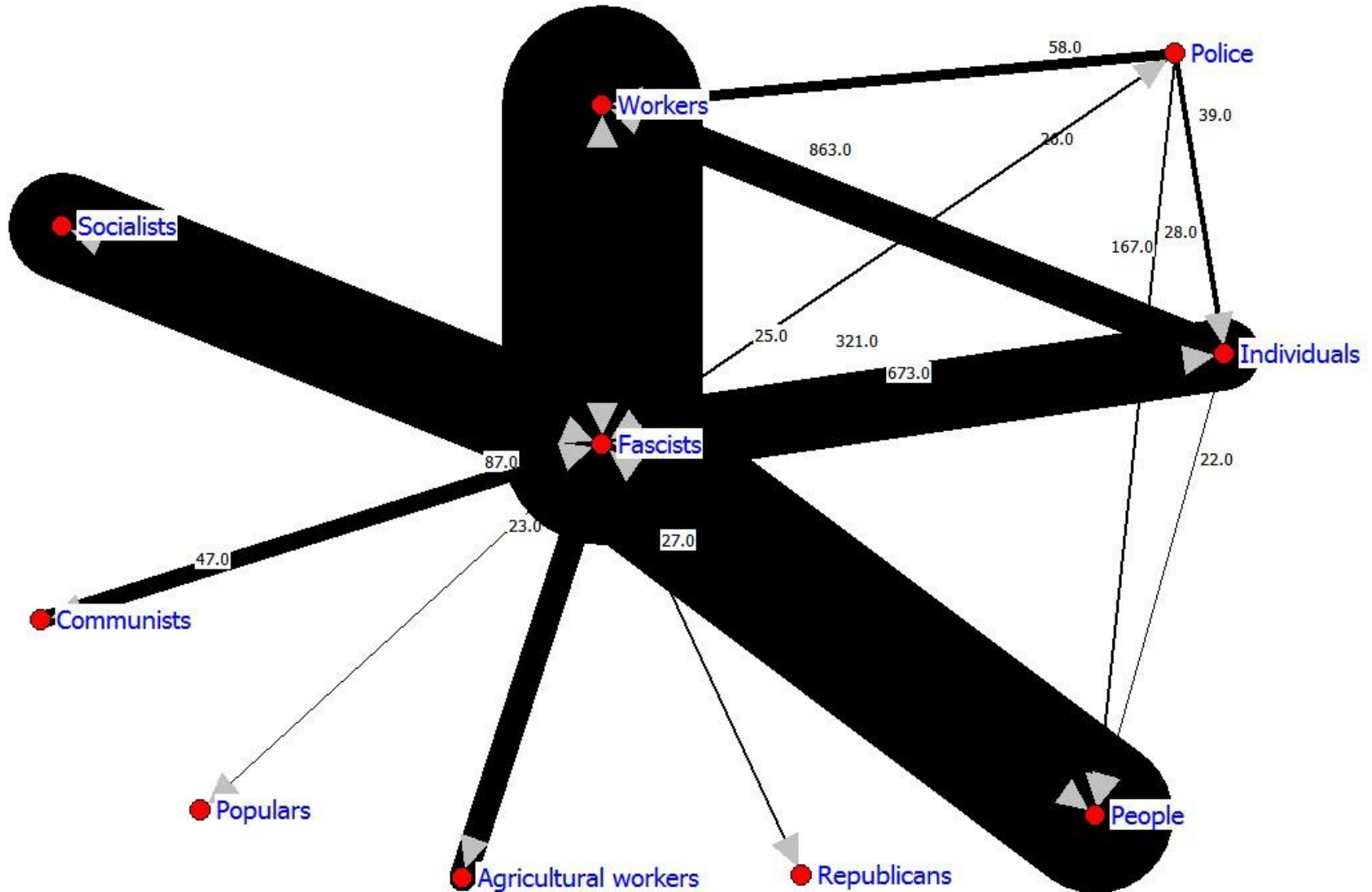
The “red years” (1919-1920)

Working-class mobilization: Police & violence

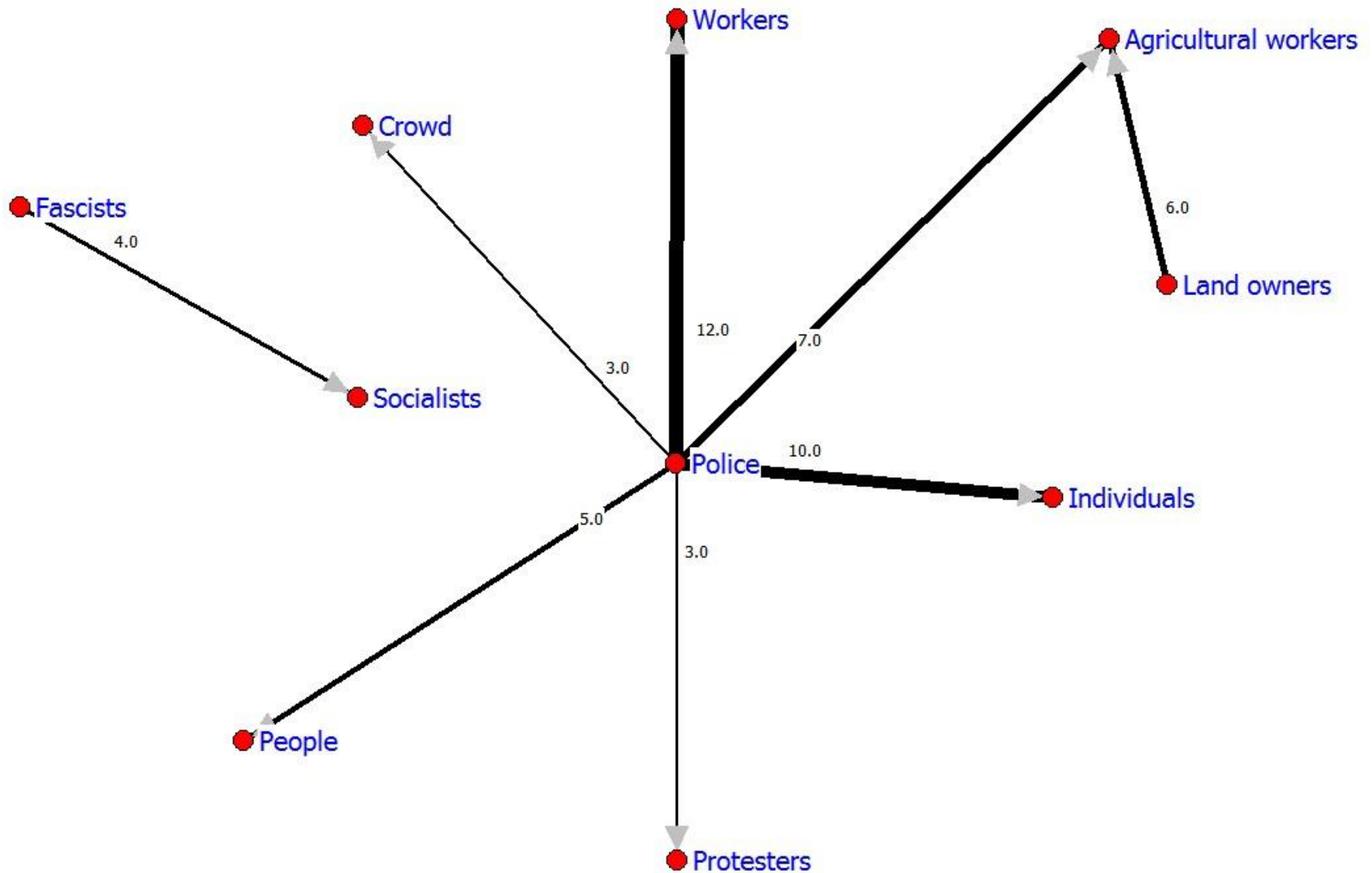


The “black years” (1921-1922)

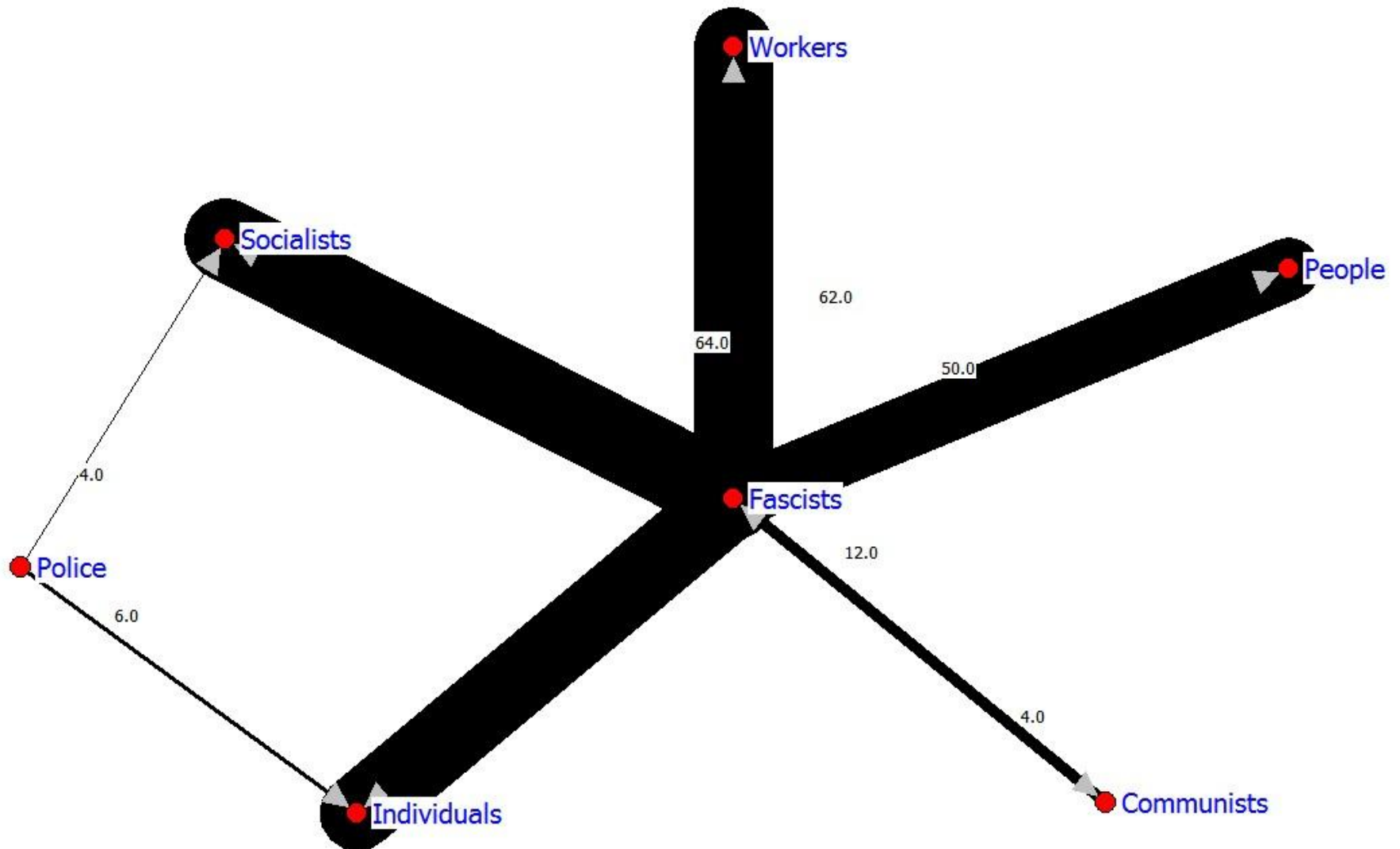
Counter mobilization: Fascists & violence



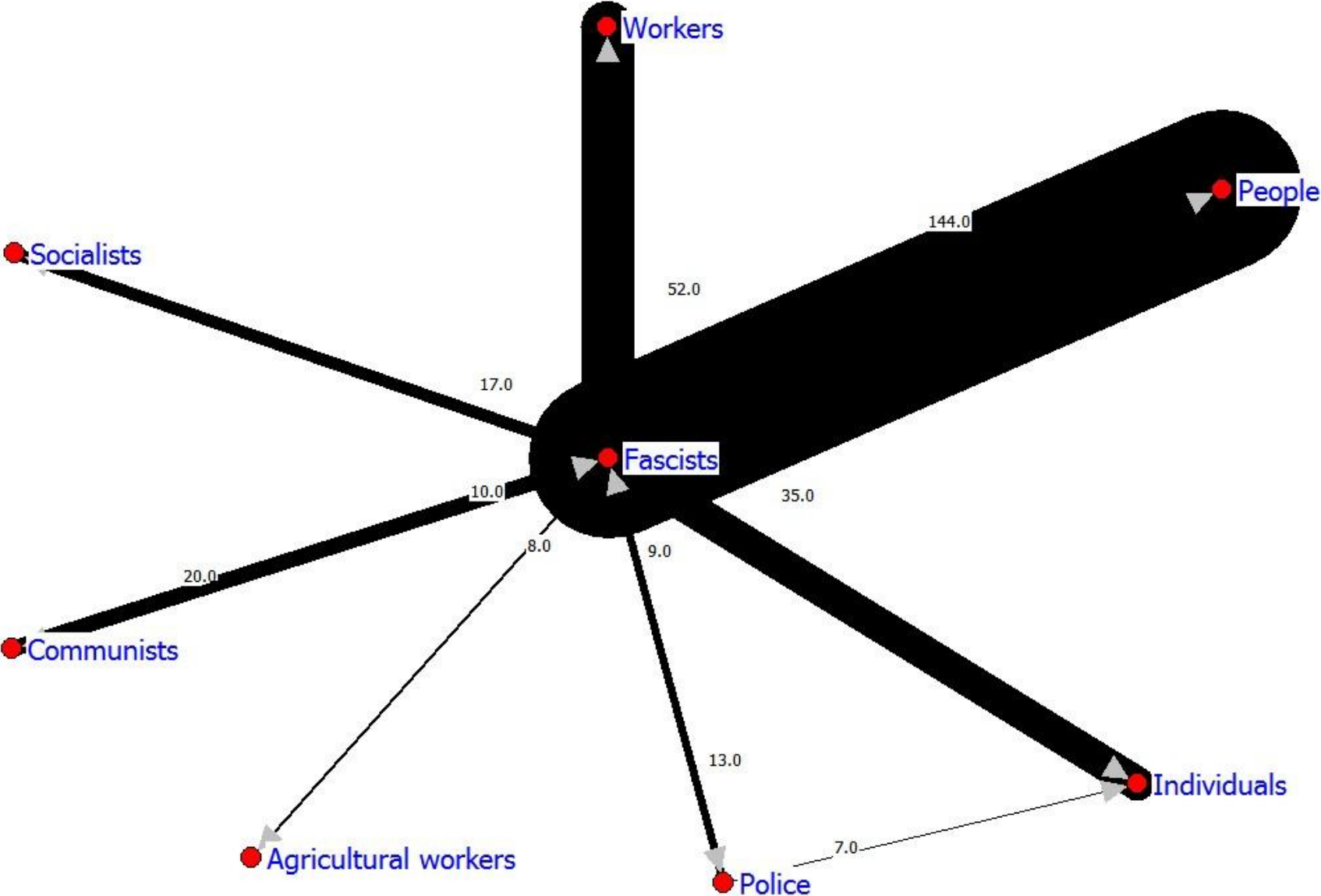
1920, 3rd Quarter (Factory Occupation)



1921, May



1921, July



Pros & Cons

- **Pros**

- Linguistic theory in coding scheme design
- Nearly hypothesis-free data
- Higher data reliability (semantic coherence)
- **Actor/action-based NOT variable-based**
fundamentally different kind of questions

- **Cons**

- Works well with narrative type of texts only
- Very labour intensive

What next?

- **In the immediate future**

- Combine RDBMS design of PC-ACE with XML (working with James Lu, Emory CS)

- **In the near future**

- Automatic solution to text parsing (working with Alberto Apostolico and Ashwin Ram, Georgia Tech CS, and Eugene Agichtein and James Lu, Emory CS)

The “most wanted” list

- News ... violence, crime, politics, ...
- Blogs
- User groups (e.g., health, computer)
- Mission statements (e.g., universities, firms)
- Personal profiles (e.g., match.com, Facebook, faculty webpages)
- Email exchanges
- Open-ended survey questions
- ...

The social scientists' approach

- Content analysis (thematic)
- CAQDAS (Atlas.ti, NVivo, MaxQda, ...)
- Automated approaches to narrative
SAO (KEDS, IDEA)

The end