Paper Recommender System

January 12, 2022

1 Load Data

[1]:
```
import pandas as pd
```

[4]:
```
df = pd.read_xml('./JASSS_DATA/jasss/jasss2020_v2.xml')
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 49 entries, 0 to 48
Data columns (total 7 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   Title         49 non-null     object
 1   Authors       49 non-null     object
 2   Institutions  49 non-null     object
 3   Issue         49 non-null     object
 4   Abstract      49 non-null     object
 5   Keywords      49 non-null     object
 6   URL           49 non-null     object
dtypes: object(7)
memory usage: 2.8+ KB
```

[5]:
```
df.head()
```

[5]:

	Title	Authors
0	Catch Me if You Can: Using a Threshold Model t…	Elizabeth A. Stiles , Colin D. Swearingen , Li…
1	How Policy Decisions Affect Refugee Journeys i…	Diana Suleimenova and Derek Groen
2	Methodological Issues of Spatial Agent-Based M…	Steven Manson , Li An, Keith C. Clarke, Alison…
3	LevelSpace: A NetLogo Extension for Multi-Leve…	Arthur Hjorth, Bryan Head, Corey Brady and Uri…
4	Cascading Impacts of Payments for Ecosystem Se…	Li An , Judy Mak , Shuang Yang, Rebecca Lewiso…

```
                                          Institutions  \
0  John Carroll University, United States; PNC Ba…
1  Brunel University London, Department of Comput…
2  Department of Geography, Environment and Socie…
3  Department of Computer Science, Aarhus Univers…
4  PKU-SDSU Complex Human-Environment Systems Cen…

                                                Issue  \
0  Journal of Artificial Societies and Social Sim…
1  Journal of Artificial Societies and Social Sim…
2  Journal of Artificial Societies and Social Sim…
3  Journal of Artificial Societies and Social Sim…
4  Journal of Artificial Societies and Social Sim…

                                             Abstract  \
0  The invisible primary is an important time in …
1  Forced displacement has a huge impact on socie…
2  Agent based modeling (ABM) is a standard tool …
3  Multi-Level Agent-Based Modeling (ML-ABM) has …
4  The theory and practice associated with paymen…

                                            Keywords  \
0  Social Network Analysis, Threshold Model, Invi…
1  Refugee Modelling, Agent-Based Modelling, Auto…
2  Spatial, Agent-Based Model, Methods, Human-Env…
3  Multi-Level, Agent-Based Modeling, Modeling To…
4  Agent-Based Modeling, Payments for Ecosystem S…

                                 URL
0  https://www.jasss.org/23/1/1.html
1  https://www.jasss.org/23/1/2.html
2  https://www.jasss.org/23/1/3.html
3  https://www.jasss.org/23/1/4.html
4  https://www.jasss.org/23/1/5.html
```

[6]:
```
df = df[['Title', 'Authors', 'Abstract', 'Keywords']]
```

[7]:
```
df.head()
```

[7]:
```
                                               Title  \
0  Catch Me if You Can: Using a Threshold Model t…
1  How Policy Decisions Affect Refugee Journeys i…
2  Methodological Issues of Spatial Agent-Based M…
3  LevelSpace: A NetLogo Extension for Multi-Leve…
4  Cascading Impacts of Payments for Ecosystem Se…

                                             Authors  \
```

```
0  Elizabeth A. Stiles , Colin D. Swearingen , Li…
1                  Diana Suleimenova and Derek Groen
2  Steven Manson , Li An, Keith C. Clarke, Alison…
3  Arthur Hjorth, Bryan Head, Corey Brady and Uri…
4  Li An , Judy Mak , Shuang Yang, Rebecca Lewiso…

                                             Abstract  \
0  The invisible primary is an important time in …
1  Forced displacement has a huge impact on socie…
2  Agent based modeling (ABM) is a standard tool …
3  Multi-Level Agent-Based Modeling (ML-ABM) has …
4  The theory and practice associated with paymen…

                                             Keywords
0  Social Network Analysis, Threshold Model, Invi…
1  Refugee Modelling, Agent-Based Modelling, Auto…
2  Spatial, Agent-Based Model, Methods, Human-Env…
3  Multi-Level, Agent-Based Modeling, Modeling To…
4  Agent-Based Modeling, Payments for Ecosystem S…
```

[8]:
```
df['Description'] = df['Abstract'] + ' ' + df['Keywords']
df.dropna(inplace=True)
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 49 entries, 0 to 48
Data columns (total 5 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   Title        49 non-null     object
 1   Authors      49 non-null     object
 2   Abstract     49 non-null     object
 3   Keywords     49 non-null     object
 4   Description  49 non-null     object
dtypes: object(5)
memory usage: 2.3+ KB
```

[9]:
```
df.head()
```

[9]:
```
                                               Title  \
0  Catch Me if You Can: Using a Threshold Model t…
1  How Policy Decisions Affect Refugee Journeys i…
2  Methodological Issues of Spatial Agent-Based M…
3  LevelSpace: A NetLogo Extension for Multi-Leve…
4  Cascading Impacts of Payments for Ecosystem Se…

                                             Authors  \
```

```
0  Elizabeth A. Stiles , Colin D. Swearingen , Li…
1                  Diana Suleimenova and Derek Groen
2  Steven Manson , Li An, Keith C. Clarke, Alison…
3  Arthur Hjorth, Bryan Head, Corey Brady and Uri…
4  Li An , Judy Mak , Shuang Yang, Rebecca Lewiso…

                                             Abstract  \
0  The invisible primary is an important time in …
1  Forced displacement has a huge impact on socie…
2  Agent based modeling (ABM) is a standard tool …
3  Multi-Level Agent-Based Modeling (ML-ABM) has …
4  The theory and practice associated with paymen…

                                             Keywords  \
0  Social Network Analysis, Threshold Model, Invi…
1  Refugee Modelling, Agent-Based Modelling, Auto…
2  Spatial, Agent-Based Model, Methods, Human-Env…
3  Multi-Level, Agent-Based Modeling, Modeling To…
4  Agent-Based Modeling, Payments for Ecosystem S…

                                          Description
0  The invisible primary is an important time in …
1  Forced displacement has a huge impact on socie…
2  Agent based modeling (ABM) is a standard tool …
3  Multi-Level Agent-Based Modeling (ML-ABM) has …
4  The theory and practice associated with paymen…
```

2 Text pre-processing

[10]:
```
import nltk
import re
import numpy as np
```

[11]:
```
stop_words = nltk.corpus.stopwords.words('english')
```

[12]:
```
def normalize_document(doc):
    # lower case and remove special characters
    doc = re.sub(r'[^a-zA-Z0-9\s]', '', doc)
    doc = doc.lower()
    doc = doc.strip()
    # tokenize document
    tokens = nltk.word_tokenize(doc)
    # filter stopwords out of document
    filtered_tokens = [token for token in tokens if token not in stop_words]
    # re-create document from filtered tokens
    return ' '.join(filtered_tokens)
```

[18]:
```
normalize_corpus = np.vectorize(normalize_document)
```

[19]:
```
norm_corpus = normalize_corpus(list(df['Description']))
```

[20]:
```
norm_corpus
```

[20]:
```
array(['invisible primary important time united states presidential primary
politics candidates gain momentum campaigns compete formally first state caucus
iowa primaries eg new hampshire however critical period possible observe hence
name simulating networks primary followers explicate hypotheses messages travel
networks affect voter preferences use threshold model drive simulated network
analysis testing spread public support candidates invisible primaries assign
voter thresholds candidates vary number voters attachment candidates decay also
vary social graph structure model results algorithm show effects size lead
unwavering base support information loss social network analysis threshold model
invisible primary campaigns elections presidency',
       'forced displacement huge impact society today 68 million people forcibly
displaced worldwide existing methods forecasting arrival migrants especially
refugees may help us better allocate humanitarian support protection however
researchers investigated effects policy decisions border closures movement
refugees recently established simulation development approaches made possible
conduct study paper use approach investigate effect policy decisions refugee
arrivals south sudan refugee crisis make study feasible terms human effort rely
agentbased modelling automated several phases simulation development using
fabflee automation toolkit observe decrease average relative difference 0615
0499 improved simulation model additional information moreover conclude border
closure reduction camp capacity induce fewer refugee arrivals time spend
travelling camps border opening increase camp capacity result limited increase
refugee arrivals destination camps best knowledge first conduct investigation
conflict refugee modelling agentbased modelling automation toolkit policy
decisions validation sensitivity analysis',
       'agent based modeling abm standard tool useful across many disciplines
despite widespread mounting interest abm even broader adoption hindered set
methodological challenges run issues around basic tools need complete conceptual
foundation approach several decades progress abms remain difficult develop use
many students scholars policy makers difficulty holds especially true models
designed represent spatial patterns processes across broad range human natural
humanenvironment systems paper describe methodological challenges facing
development use spatial abm sabm suggest potential solutions multiple
disciplines first define sabm narrow object inquiry explore spatiality source
advantages challenges examine time interacts space models delve issues model
development general modeling frameworks tools specifically draw lessons insights
fields history abm contributions including economics ecology geography ecology
anthropology spatial science goal identifying promising ways forward powerful
means modeling spatial agentbased model methods humanenvironment systems',
       'multilevel agentbased modeling mlabm receiving increasing attention
recent years paper present levelspace extension allows modelers easily build
```

mlabms popular widely used netlogo language present levelspace framework associated programming primitives based three common usecases mlabm coupling heterogenous models dynamic adaptation detail crosslevel interaction show easy build mlabms levelspace argue important unified conceptual language describing levelspace models present six dimensions along models differ discuss combined variety mlabm types levelspace finally argue future work explore relationships six dimensions different configurations might less appropriate particular modeling tasks multilevel agentbased modeling modeling tools netlogo',

'theory practice associated payments ecosystem services pes feature variety piecemeal studies related impacts socioeconomic demographic environmental variables lacking efforts understanding mutual relationships spatially temporally explicit manner addition pes literature short ecological metrics document consequences pes land use land cover change building detailed survey data fanjingshan national nature reserve fnnr china developed tested agentbased model study complex interactions among human livelihoods migration resource extraction particular pes guizhou golden monkey habitat occupancy 20 years performed simulationbased experiments testing social ecological impacts pes payments well human population pressures results show steady increase outmigration number land parcels enrolled one chinas major pes programs tends increase reach peak slowly decline showing convex trend converges stable number enrolled parcels regardless payment levels simulated monkey occupancy responds changes pes payment levels substantially edge areas fnnr model useful fnnr also applicable platform study understand human ecological roles pes many complex humanenvironment systems shedding light key elements interactions relationships systems pes researchers practitioners bear mind research contributes establishing scientific basis pes science incorporates features complex systems offering realistic spatially temporally explicit insights related pes policy related interventions agentbased modeling payments ecosystem services complex humanenvironment systems guizhou snubnosed monkey migration land use',

'designing implementing applying agentbased models abms requires structured approach part comprehensive analysis output input variability form uncertainty sensitivity analysis sa objective paper assist choosing given abm appropriate methods sa argue single sa method fits abms different methods sa used based overarching purpose model example abstract exploratory models focus deeper understanding target system properties fed critical data representing patterns stylized facts simple sa methods may sufficient capturing dependencies outputinput spaces contrast applied models used scenario policyanalysis usually complex datarich higher level realism required choice sophisticated sa may critical establishing robustness results model results passed endusers accordingly present roadmap guides abm developers process performing sa best fits purpose abm roadmap covers wide range abm applications advocates routine use global methods capture input interactions therefore mandatory scientists want recognize sensitivities part roadmap report frontier sa methods emerging recent years handling temporal spatial outputs b using whole output distribution result rather variance c looking topological relationships input data points rather values looking abm black box finding behavioral primitives using study complex system characteristics like regime shifts tipping points condensation

versus dissipation collective system behavior sensitivity analysis agentbased model individualbased model review',

'paper introduce simple way parametrize simulation models using regularized linear regression regressions bypass three major challenges calibrating minimization selecting summary statistics defining distance function minimizing numerically substituting regression classification extend approach model selection present five example estimations statistical fit biological individualbased model simple real business cycle model nonlinear biological simulation heuristics selection fishery agentbased model outcome method automatically chooses summary statistics weighs uses parametrize models without running direct minimization agentbased models indirect inference estimation calibration simulated minimum distance approximate bayesian computation',

'drivers conversion organic farming still residual choice agriculture poorly understood many scholars argue farm characteristics determine choice exclude social dynamics study issue developed agentbased model agent decisions shift organic farming based comparison satisfaction current situation potential satisfaction alternative farming strategy farmer agents satisfaction modelled theory reasoned action implied compare agent productions time agents former lend great credibility important others moreover famer operates technical changes affecting productions imitating credible farmers used model examine simple abstract farm populations adapted agricultural census data farm characteristics dairy farming 27 french cantons exploiting domain expertise data previous research proposed laws model impact conversion terms milk environmental amenities productions farm simulations real populations farms confirm strong impact farm characteristics however results also suggest complex impact social dynamics favour disfavour diffusion organic farming dynamic implicit networks similarity credibility confirmed strong importance demographic evolution farm characteristics organic farming adaptation credibility theory reasoned action agentbased model social influence',

'ictbased collaborative innovation significant impact economy facilitating technological convergence promoting innovation industries however research innovation suggests polarization firm size distribution grown since early 2000s interfere collaborative innovation among firms paper modelled firms decisionmaking processes led collaborative innovation spatial nperson iterated prisoners dilemma nipd game using collaborative innovation data korean ict firms using agentbased model experimented effects firm size heterogeneity collaborative innovation simulation experiment results reveal collaborative innovation industry increases size heterogeneity decreases findings suggest policies promoting collaborative innovation focus mitigating structural inequalities industry agentbased modelling prisoners dilemma pavlovian cooperation collaborative innovation firm size distribution ict industry',

'modelling simulation play increasingly significant role exploratory studies informing policy makers climate change mitigation strategies considerable research done creating integrated assessment models iams focus examining human impacts climate change many popular iams created steady state optimisation models typically employ nested structure neoclassical production functions represent energyeconomy system holding aggregate views variables hence

unable capture finer level details underlying system components alternative approach allows modelling populations collection individual unevenly distributed entities agentbased modelling often used field social simulation simulating huge numbers individual entities quickly become issue requires large amounts computational resources goal paper introduce conceptual framework developing hybrid iams novel modelling approach allows us reuse existing rigid wellestablished iams adds flexibility replacing aggregate stocks community vibrant interacting entities provide proofofconcept application conceptual framework form illustrative example test case takes settings us solely created purpose demonstrating hybrid modelling approach claim predictive powers integrated assessment modelling climate change agentbased modelling system dynamics modelling methodological advance hybridisation scalability',

'international agricultural trade changed land uses trading countries altering global food security environmental sustainability studies concluded local landuse drivers largely global sources eg trade increases deforestation exporting countries however little known local landuse changes affect distant locations namely feedback yet distant impacts feedbacks significant governing local land systems framework telecoupling ie socioeconomicenvironmental interactions distant places shown effective conceptual tool study international trade associated socioeconomic environmental impacts however systems simulation tool quantify telecoupled causes effects still lacking construct new type agentbased model abm simulate landuse changes multiple distant places namely teleabm telecoupled agentbased model use soybean trade brazil china example brazil sending system china receiving system worlds largest soybean exporter importer respectively select one representative county country calibrate validate model spatiotemporal analysis historical landuse changes empirical analysis household survey data describe model following oddd protocol validate model results location respectively illustrate aggregated farmer agents landuse behaviors sending system result landuse changes receiving system vice versa one scenario example ie hightariff scenario given demonstrate results teleabm model allows us advance understanding telecoupling features influence land system science test hypotheses complex coupled humannatural systems eg cascading effect telecoupling agentbased model land system landuse change soybean trade oddd',

'using agentbased model miller et al 2012 depicts different types individuals memory affect formation performance organizational routines show replicated simulation model used develop theory also assess standards odd overview design concepts details protocol doe design experiments principles support replication evaluation analysis model using verified model conduct several simulation experiments examples different types theory development first show previous theoretical insights generalized investigating additional scenarios mergers second show potential replicated simulation models theory refinement analyzing indepth relationship memory functions routine performance routine adaptation replication abm odd design experiments doe organizational routines dynamic capabilities',

'editorial paper reviews state science agentbased modeling abm pointing strengths weaknesses abm paper also highlights several impending tasks warrant special attention order improve science application abm modeling human decisions

abm transparency reusability validation abm abm software big data abm abm theories six innovative papers included special issue summarized connections abm impending tasks brought attention authors hope special issue help prioritize specific resources activities relation abm advances leading coordinated joint efforts initiatives advance science technology behind abm agentbased modeling complex systems system integration socialecological systems overview',

'dynamics social stigma explored context diffusion models focus exploring dynamic process behavior individuals interpersonal relationships among influence macrosocial attitude towards stigma find norm tolerance best promoted population comprises whose conduct driven compassion stigmatized whose focus conforming others social networks second finding less insular social networks encourage destigmatization people compassionate instead insularity promotes tolerance society dominated conformity stigma diffusion conformity compassion social network',

'managing disasters caused natural events especially volcanic crises requires range approaches including risk modelling analysis risk modelling commonly conducted communityregional scale using gis however people objects move response crisis static approaches capture dynamics risk properly accommodate objects movements within time space emergence agentbased modelling makes possible model risk individual level evolves space time propose new approach spatiotemporal dynamics model risk stdmr integrating multicriteria evaluation mce within georeferenced agentbased model using mt merapi indonesia case study model makes possible simulate spatiotemporal dynamics risk volcanic crisis importantly individual vulnerability heterogeneous depends characteristics individuals concerned risk individuals dynamic changes along hazard location model able highlight small number highrisk spatiotemporal positions due behaviour individuals evacuating volcano dynamics hazard overall risk times places extremely high outcomes extremely relevant stakeholders work coupling abm mce dynamic volcanic hazard novel contextually relevant abm volcanic crisis risk estimation spatiotemporal modeling mce merapi',

'human interactions opinion exchanges lead social opinion dynamics well described opinion formation models models random parameter usually considered system noise indicating individuals inexplicable opinion changes noise could indicator influential factors public media affects emotions study phase transitions changes one social phase another various noise levels discrete opinion formation model based social impact theory scalefree random network interaction network topology also generate another similar model using concept social power based agents node degrees interaction network estimation persuasiveness supportiveness strengths compare models phase transition viewpoint show agentbased simulation analytical considerations opinion phases including majority nonmajority formed terms initial population agents opinion groups noise levels two factors affect system phase equilibrium noise level increases breaking segregated groups dominance stochastic behavior agents deterministic behavior high enough noise levels system reaches nonmajority phase equilibrium regardless initial combination opinion groups relatively low noise levels original model model whose agents strengths proportional centrality different behaviors presence highconnected influential leaders latter model

consequences different behavior reaching equilibrium phase different thresholds noise levels phase transitions opinion formation noise agentbased modeling social impact model phase transition',

'drivers conversion organic farming still residual choice agriculture poorly understood many scholars argue farm characteristics determine choice exclude role social dynamics study issue developed agentbased model agents decisions shift organic farming based comparison satisfaction current situation potential satisfaction alternative farming strategy farmer agents satisfaction modelled using theory reasoned action makes necessary compare agents productions time agents former attributes considerable credibility important others moreover farmers make technical changes affect productions imitating credible farmers first used model examine simple abstract farm populations also adapted use data agricultural census concerning farm characteristics dairy farming 27 french cantons based domain expertise data previous research propose certain laws modelling impact conversion farm production milk environment simulations real populations farms confirm important impact farm characteristics however results also suggest complex impact social dynamics favour impede diffusion organic farming dynamic implicit networks similarity credibility confirm great importance demographic changes farm characteristics organic farming adaptation theory reasoned action agentbased model social influence credibility',

'present model showing evolution organization agents discuss democratically good practices model feeds field study twelve years france followed npos called amap observed construction time regional national level hypothesis make either based literature opinion diffusion results field study defining dynamics agents influence make collective decision group level decide stay leave respective groups analyse effect different forms vertical communication meant spread good practices within organization main indicators good functioning democratic dynamics stability representativeness show communication norms well designed positive impact stability representativeness interestingly effect communication increases number dimensions discussed groups communication norms thus valuable tool use groups wish improve democratic practices without jeopardizing stability agentbased model communication opinion dynamics democracy nonprofit organization short food chain',

'realworld social networks often exhibit high levels clustering positive degree assortativity short average path lengths smallworld property rightskewed rarely power law degree distributions hand homophily defined propensity similar agents connect one fundamental social processes observed many human animal societies paper examine extent homophily sufficient produce typical structural properties social networks conduct simulation study based social distance attachment sda model particular kind random geometric graph rgg nodes embedded social space connection probabilities depend functionally distances nodes derive form model first principles based existing analytical results argue mathematical construction rggs corresponds directly homophily principle provide good model find homophily especially combined random edge rewiring sufficient reproduce many characteristic features social networks additionally devise hybrid model combining sda configuration model allows generating homophilic networks arbitrary degree sequences use study interactions homophily processes imposing

constraints degree distributions show effects homophily clustering robust respect distribution constraints degree assortativity highly dependent particular kind enforced degree sequence social networks homophily social distance attachment configuration model',

'overview design concepts details odd protocol describing individual agentbased models abms widely accepted used document models journal articles standardized document providing consistent logical readable account structure dynamics abms research groups also find useful workflow model design even still limitations odd obstruct widespread adoption limitations discussed addressed paper limited availability guidance use odd length odd documents limitations odd highly complex models lack sufficient details many odds enable reimplementation without access model code lack provision sections document structure covering model design rationale models underlying narrative means models fitness purpose evaluated document steps taken provide better guidance structuring complex odds odd summary inclusion journal article full details supplementary material table 1 using odd point readers relevant sections model code update document structure include sections model rationale evaluation also advocate need standard descriptions simulation experiments argue odd principle used type simulation model thereby odd would provide lingua franca simulation modelling agentbased model individualbased model best practice simulation model standardization documentation',

'study impact endogenous creation destruction social ties artificial society aggregate outcomes generalized trust willingness cooperate social utility economic performance end put forward computational multiagent model agents overlapping generations interact dynamically evolving social network model four distinct dimensions individuals social capital degree centrality heterophilous homophilous interactions determine generalized trust willingness cooperate altogether helping achieve certain levels social utility ie utility social contacts economic performance find stationary state simulated social network exhibits realistic smallworld topology also observe societies whose social networks relatively frequently reconfigured display relatively higher generalized trust willingness cooperate economic performance cost lower social utility similar outcomes found societies social tie dissolution relatively weakly linked family closeness social network structure social network dynamics trust willingness cooperate economic performance agentbased model',

'recent advancement agentbased modeling characterized higher demands parameterization evaluation documentation computationally expensive models accordingly also growing request easy go applications mimicking inputoutput behavior models metamodels increasingly used tasks paper provide overview common metamodel types purposes usage agentbased modeling context guide modelers selection application metamodels needs assessed implementation effort performance performed literature research january 2019 using four different databases five different terms paraphrasing metamodels approximation emulator metamodel metamodel surrogate used capture whole range relevant literature disciplines metamodel applications found categorized specific metamodel types rated different junior senior researches varying disciplines including forest sciences landscape ecology economics regarding implementation effort performance

specifically captured metamodel performance according consideration uncertainties ii suitability assessment provided authors particular purpose iii number valuation criteria provided suitability assessment selected 40 distinct metamodel applications studies published peerreviewed journals 2005 2019 used sensitivity analysis calibration upscaling agentbased models well mimic prediction different scenarios review provides information applicable metamodel types purpose forms first guidance implementation validation metamodels agentbased models individualbased model surrogate model emulator calibration sensitivity analysis review',

'covid19 pandemic causing dramatic loss lives worldwide challenging sustainability health care systems threatening economic meltdown putting pressure mental health individuals due social distancing lockdown measures pandemic also posing severe challenges scientific community scholars pressure respond policymakers demands advice despite absence adequate trusted data understanding pandemic requires finegrained data representing specific local conditions social reactions individuals experts built simulation models estimate disease trajectories may enough guide decisionmakers formulate policy measures limit epidemic cover full behavioural social complexity societies pandemic crisis modelling large potential impact upon peoples lives great responsibility paper calls scientific community improve transparency access rigour models also calls stakeholders improve rapidity data trusted sources released community fully responsible manner responding pandemic stress test collaborative capacity socialeconomic value research covid19 pandemic disease agentbased models modelling policy data',

'paper introduces mbssm mechanismbased social systems modelling software architecture designed expressing mechanisms social theories individual behaviour components unified way implementing mechanisms agentbased simulation model mbssm architecture based middlerange theory approach recently expounded analytical sociology designed objectoriented programming paradigm unified modelling language diagrams paper presents two worked examples using architecture modelling individual behaviour mechanisms give rise dynamics populationlevel alcohol use singletheory model norm theory multitheory model combines norm theory role theory mbssm architecture provides computational environment within theories based social mechanisms represented compared integrated architecture plays fundamental enabling role within wider simulation modelbased framework abductive reasoning families theories tested ability explain concrete social phenomena agentbased modelling social simulation software architecture analytical sociology abductive reasoning',

'blockchain viewed public ledger maintained collectively large number participators based consensus protocol interested difference consensus protocols trade network topologies affect performance blockchain system studied literature yet paper proposed agentbased model consisting multiple trader miner agents one system agent investigated three consensus protocols namely proofofwork pow proofofstake pos delegated proofofstake dpos also examined three common trade network topologies random smallworld scalefree find consensus protocol trade network topology impact performance blockchain system pos dpos generally better pow terms increasing trade efficiency equalizing wealth besides scalefree trade

network favorable trade efficiency quite low moderates price fluctuation wealth inequality since connectivity inequality determines wealth inequality crucial increase connectivity among participants designing sustainable blockchain system suggest findings could useful designers practitioner researchers blockchain system token economy blockchain consensus protocol trade network topology agentbased model',

'agentbased modelling valuable approach modelling systems whose behaviour driven interactions distinct entities crowds people however faces fundamental difficulty established mechanisms dynamically incorporating realtime data models limits simulations inherently dynamic pedestrian movements scenario testing historic patterns rather realtime simulation present paper demonstrates particle filter could used incorporate data agentbased model pedestrian movements run time experiments show although possible use particle filter perform online real time model optimisation number individual particles required hence computational complexity increases exponentially number agents furthermore paper assumes onetoone mapping observations individual agents would case reality therefore paper lays fundamental groundwork highlights key challenges need addressed realtime simulation crowd movements become reality success could implications management complex environments nationally internationally transportation hubs hospitals shopping centres etc agentbased modelling particle filter data assimilation crowd simulation pedestrian modelling',

'people tend form groups solve difficult problems groups seem better problemsolving capabilities individuals indeed evolution human beings learned cooperation frequently optimal strategy solve hard problems quickly accurately ability group determine solution given problem group members alone called collective intelligence emergent property group whole result complex interaction many factors propose simple analytically solvable model disentangling direct link collective intelligence average intelligence group members found nonlinear relation collective intelligence group average intelligence quotient members depending task difficulty found three regimes follows simple tasks level collective intelligence group decreasing function teammates intelligence quotient tasks intermediate difficulties relation collective intelligence intelligence quotient shows nonmonotone behaviour complex tasks level collective intelligence group monotonically increases teammates intelligence quotient phase transitions emerging varying latters level although simple abstract model paves way future experimental explorations link task complexity individual intelligence group performance collective intelligence problem solving emergent behaviour',

'polarization threatening stability democratic societies polarization research focused opinion extremeness overlooking correlation different policy issues paper explain emergence hyperpolarization ie combination extremeness correlation issues developing new theory opinion formation called weighted balance theory wbt wbt extends heiders cognitive balance theory encompass multiple weighted attitudes validated wbt empirical data 2016 national election survey furthermore developed opinion dynamics model based wbt first time able generate hyperpolarization explain link affective opinion polarization finally theory encompasses phenomena opinion dynamics including monopolarization

backfire effects polarization balance theory opinion dynamics agentbased modeling',

'recently financial markets shown significant risks levels volatility understanding sources risks require simulation models capable representing adequately real mechanisms markets paper compared data highfrequencytrader marketmaking hftmm strategy real financial market simulation regarding former extracted trader clusters identified one cluster whose statistical indexes indicated hftmm features analyzed difference traders orders market price simulation built artificial market model continuous double auction system stylized trader agents hftmm trader agents based prior research experiment compared distribution order placements hftmm traders real simulated financial data found order placement distribution near market best price real data simulations similar however orders far market best price differed significantly real data exhibited wider range orders indicates order build realistic simulation financial markets integrating finegrained data essential artificial market multiagent simulation datamining highfrequency trade marketmaking clustering',

'ongoing discussion concerning relationship social welfare climate change thus required level type measures needed protect climate integrated assessment models iams extended incorporate technological progress heterogeneity uncertainty making use stochastic dynamic equilibrium approach order derive solution according literature iam class models take relationships among economic social environmental factors account moreover consider interdependencies microlevel meaning possible consequences duly examined propose agentbased approach analyse relationship economic welfare climate protection particular aim analyse decisions individual agents allowing tradeoff economic welfare climate protection influence aggregated emergent economic behaviour using model estimate damage function values order 3 4for 2 c temperature increase linear slightly concave shape show heterogeneity agents technological progress damage function may lead lower gdp growth rates greater temperaturerelated damage forecast models solely homogeneous representative agents climate change climate protection integrated assessment model agentbased modelling',

'simulation models proven valuable tools studying peer review processes however effects models assumptions tested models examined comparative contexts paper address two assumptions go tandem 1 granularity evaluation scale 2 homogeneity grade language ie whether reviewers interpret evaluation grades fashion test consequences assumptions extending wellknown agentbased model author reviewer behaviour discrete evaluation scales reviewers interpretation grade language way compare peer review model homogeneous grade language assumed models peer review psychologically realistic model reviewers interpret grades evaluation scale heterogeneously find grade language heterogeneity indeed affect predictions model peer review peer review grade language agentbased modeling',

'reflexive phenomena usually understood social sciences processes affect recursively stems mutual altering relationship participants social process belong participants change course process actions new state evolution process lead change participants behavior article proposes agentbased model diffusion innovations social network study reflexivity model agents decide adopt new

product according utility function depends two kinds social influences first local influence exerted agent closest neighbors already adopted also feels product suits personal needs second global influence leads agents adopt become aware emerging trends happening system endow agents reflexive capacity allows recognize trend even perceive significant change neighborhood results reveal appearance slowdown periods along adoption rate curve contrast classic stylized bellshaped behavior results also show network structure plays important role effect reflexivity structures eg scalefree networks may amplify others eg smallworld structure weaken effect contribution work lies inclusion evolving cognitive distinctions agents decide product adoption diffusion processes reflexivity diffusion innovations secondorder emergence global network externalities',

'social computational archaeology focuses largely study past societies evolution human behaviour time agentbased models abms allow efficient modeling human agency quantitative representation exploration specific properties patterns archaeological information work put forward novel agentbased trading model simulating exchange distribution resources across settlements past societies model part broader abm populated autonomous utilityseeking agents corresponding households ability employ spatial interaction model choice allows study settlements trading ability power given geolocation position within trading network structural properties network case study use minoan society bronze age wider area knossos island crete greece instantiate two wellknown spatial interaction submodels xtent gravity conduct systematic evaluation dynamic trading network formed time simulations assess sustainability artificial minoan society terms population size number distribution agent communities respect available archaeological data spatial interaction model employed evaluate resulting trading networks structure centrality clustering etc affects intersettlement organization providing process insights support archaeological hypotheses settlement organization place time results show trading network modeled using gravity focuses settlements importance rather proximity settlement numbers evolution patterns emerge similar ones exist archaeological record also inferred simulations rather dense trading network without strict settlement hierarchy could emerged late minoan period theran volcanic eruption well documented historic catastrophic event moreover appears trading networks structure interaction patterns reversed theran eruption compared effect earlier periods agentbased modeling modelbased archaeology spatial interaction model graphtheory trade network minoan civilization',

'drives prices fine wines easy discern view multitude confounding factors characterising transactions across several markets time understanding quantitative relationships mechanisms determine price level important policy making eg predicting outcomes regulations methodological purposes elements consider modelling markets examine price formation fine wines simultaneously across three markets automated electronic exchange livex intermediated auctions overthecounter otc use unique dataset consisting 99769 price data points premier cru bordeaux fine wines examine price determinants bayesian modelling ascertain mean price ranking otc market expensive livex least differing 45 08 auctions also find slight price decrease larger transactions approx03 reduction 10 volume

increase platykurtosis price distribution greatest livex observe stochastic noise auctions agentbased simulation discover necessary include trading mechanisms commissions otc market heterogeneity reproduce observed characteristics results indicate elements included future fine wine markets models parallel trading trading systems price formation wine investment agent based modelling',

'paper presents simulation model describing radicalisation process radicalisation process complex human socioenvironmental process much academic interest past two decades despite still poorly understood extremely difficult area social scientists research subject suffers lack available data making construction effective simulation model particularly challenging order construct simulation paper rely theoretical framework originally developed means synthesising academic literature radicalisation theoretical framework three levels individual vulnerability radicalisation exposure radicalising moral contexts emergence radicalising settings adapt framework simulation model first reconstructing individuallevel statetransition model next appropriate data sought parameterise model parallel drawn process radicalisation process people develop propensity participate general acts criminality analogy enables considerably data used parameterisation model calibrated considering logical differences crime terrorism might lead differences radicalisation criminality development processes model validated stylised facts demonstrating despite highly theoretical simulation capable producing realistic output possible uses model evaluate effectiveness counterradicalisation measures also considered radicalisation socialecological modelling statetransition modelling model development stylized facts',

'paper describes application largescale active learning method characterize parameter space computational agentbased model developed investigate impact communityrx clinical informationbased health intervention provides patients personalized information local community resources meet basic selfcare needs diffusion information community resources use modeled via networked interactions subsequent effect agents use community resources across urban population random forest model iteratively fitted model evaluations characterize model parameter space respect observed empirical data demonstrate feasibility using highperformance computing active learning model exploration techniques characterize large parameter spaces partitioning parameter space potentially viable nonviable regions rule regions space simulation output implausible observed empirical data argue methods necessary enable model exploration complex computational models incorporate increasingly available microlevel behavior data provide public access model highperformance computing experimentation code agentbased modeling model exploration highperformance computing active learning',

'question acts selflessness occur hobbesian selfhelp world fascinated scholars decades centuries utilizing simulations previous research shown altruism evolutionarily stable smallscale societies narrow set circumstances however expanding models populations anything larger hundred people generally break paper modify widely used imagescore mechanism include contagionbased reputation demonstrate altruism survive populations 20000 also find selflessness

strongly depends network topology heavily clustered smallworld societies resemble tightknit family friendship structures promote cooperation random networks connections superficial altruism evolution network simulation smallworld',

'process beliefs opinions individual socially malleable attributes spread across society known cultural dissemination broadly recognized concept among sociologists political scientists yet fundamental aspects process ultimately lead cultural divergences rural urban segments society currently poorly understood article uses agentbased model isolate analyze one basic yet essential facet issue namely question intrinsic differences urban rural population densities influence levels cultural homogeneityheterogeneity emerge within region urban rural cultures develop isolation one another dynamical interplay two particular import evolution found urban areas relatively high number local neighbors one interact tends promote cultural homogeneity urban rural regions moreover rather surprisingly higher frequency potential interactions neighbors within urban regions promotes homogeneity urban regions tends drive rural regions towards greater levels heterogeneity cultural evolution cultural transmission opinion dynamics agentbased modeling cultural dissemination',

'paper provide overview worksim model agentbased framework designed study labor markets first objective model reproduce within rigorous stockflow accounting gross flows individuals important workstates ie employment distinguishing fixed term contracts openended contracts unemployment inactivity french legal institutions labor market modelled detail constrain decisions agents job flows worker flows firms individuals heterogeneous decisions taken basis bounded rationality yet employers well workers form imperfect anticipations one important theoretical novelty model consider multijob firms shocks individual demand firms employers consider anticipated shocks decide types contract model calibrated secondary objective characterize nature labor market study notably differentiated roles two types contracts impact unemployment achieved first examining patterns flows stocks labor secondly sensitivity experiments modifying certain exogenous parameters variables total demand used model tool experimenting labor market policies including changes labor law france agentbased simulation dual labor markets anticipations bounded rationality policy evaluation',

'paper introduces agentbased model housing market macroprudential policy experiments specifically simulation model used examine effects policy setting loantovalue ltv debttoincome dti policy instruments several governments use regulate housing market simulation model illustrates interactions among households house suppliers real estate brokers model household population either seller buyer households may behave speculators housing market better understand impact policies used realworld observations korean housing market include various economic conditions policy variables korean census data baseline model quantitatively validated price index transaction volume past korean housing market validation show empirical effectiveness setting ltv dti towards house prices transaction volumes amount households mortgages furthermore investigate simulation results owneroccupier rate households investigations provide policy analyses koreas housing market governments ltv dti regulations housing market

macroprudential policy loantovalue debttoincome agentbased modeling policy impact analysis',

'used value sensitive design method develop agentbased model values humanitarian logistics refugees schwartzs theory universal values implemented model way agents make value tradeoffs operationalized measure refugee wellbeing measure public opinion refugee logistics handled trying different valuescenarios stakeholders responsible involved refugee logistics insights effects various value choices model visualized made usable platform interactive website decisionmakers understand tradeoffs policies government nongovernment organizations agent based model value sensitive design simulation policy humanitarian logistics refugees schwartz values',

'highfidelity models increasingly used predict guide decision making prior work emphasized importance replication ensuring reliable modeling yielded important replication strategies however work based relatively simple theory generating models lessons might translate highfidelity models used decision support using netlogo replicate recently published highfidelity model examining effects hiv biomedical intervention use modular approach build model ground provide examples replication process investigating replication two submodules well overall simulation experiment first module achieved numerical identity replication whereas obtained distributional equivalence replicating second module achieved relational equivalence among overall model behaviors 098 correlation across two implementations outcome measure even without strictly following original model formation sexual network results show replication highfidelity models feasible following set systematic strategies leverage modularity highlight role replication standards modular testing functional code facilitating strategies replication agentbased models modular highfidelity hiv',

'lifting social restrictions one critical decisions public health authorities face pandemic covid19 work focuses risk associated decision called period reopening decision epidemic expiration final epidemic phase considered critical epidemic conditions could possibly emerge phase factors considered include proportion asymptomatic cases mitigation strategy based testing average duration infectious states assuming hypothetical configurations time reopening decision partial knowledge concerning epidemic dynamics available public health authorities analyzed risk reopening decision based possibly unreliable estimates presented discretetime stochastic model statedependent transmission probabilities multiagent simulations results show different outcomes produced different proportions undetected asymptomatic cases different probabilities asymptomatic cases detected contained multivariate analysis risk based average duration asymptomatic contained states finally analysis highlights enduring uncertainty typical pandemic requires risk analysis approach complement epidemiological studies stochastic epidemic model multiagent simulation network analysis agentbased model risk analysis',

'successful adoption innovations depends provision adequate information farmers rural areas developing countries farmers usually rely social networks information source hence policymakers programimplementers benefit social diffusion processes effectively disseminate information study aims identify set farmers initially obtain information seeds optimises diffusion network

systematically evaluates different criteria seed selection number seeds interaction effects empirical agentbased model adjusted case study rural zambia applied predict diffusion outcomes varying seed sets ex ante simulations revealed informing farmers connections leads highest diffusion speed reach also targeting village heads farmers high betweenness centrality function bridges connecting different parts network enhances diffusion increased number seeds improves reach marginal effects additional seeds decline interdependencies seed set size selection criteria highlight importance considering seed selection criteria seed set size optimising seeding strategies enhance information diffusion information diffusion social networks agentbased modelling seeding zambia',

'network scientists proposed infectious diseases involving persontoperson transmission could effectively halted interventions targeting minority highly connected individuals could strategy effective combating virus partly transmitted closerange contact many believe sarscov2 effectiveness critically depends high betweenperson variability number closerange contacts analyzed population survey data showing distribution closerange contacts across individuals indeed characterized small proportion individuals reporting high frequency contacts strikingly found average duration contact mostly invariant number contacts reinforcing criticality hubs simulated population embedded network empirically observed contact frequencies simulations showed targeting hubs robustly improves containment agentbased computational models complex social networks virus diffusion immunization strategies epidemiological models',

'studies colonization processes past human societies often use standard population model population represented single quantity real populations processes however structured internal classes stages classes sometimes created based social differentiation present work information colonization old providence island used create agentbased model colonization process heterogeneous environment population social differentiation agents socially divided two classes modeled dissimilar spatial clustering preferences model simulations assessed importance gregarious behavior colonization processes conducted heterogeneous environments sociallydifferentiated populations results suggest conditions colonization process starts agent cluster largest suitable area spatial distribution agents maintained tendency toward randomness simulation time increased even gregariousness values increased conspicuous effects agent clustering produced initial conditions behavioral adaptations increased agent capacity access resources likelihood gregariousness approach presented could used analyze past human colonization events support longterm conceptual design future human colonization processes small social formations unfamiliar uninhabited environments human colonization gregarious behavior social differentiation settlement patterns caribbean archaeology',

'last years use agentbased simulations study social systems spread many domains eg geography ecology sociology economy simulations aim reproduce real life situations involving human beings thus need integrate complex agents match behavior simulated people therefore notions cognition emotions personality social relationships norms taken account currently agent architecture could incorporate features used majority modelers including low levels skills

```
programming paper ben behavior emotions norms architecture introduced tackle
issue modular architecture based bdi model cognition featuring modules add
emotions emotional contagion personality social relationships norms agent
behavior architecture integrated gama simulation platform application ben
simulation evacuation nightclub fire presented shows complexity behaviors may
developed architecture create credible expressive simulations social simulation
agent architecture bdi emotions personality emotional contagion',
       'housing sector important part every community directly affects people
constitutes major share building market shapes community meanwhile increase
developments hazardprone areas along intensification extreme events amplified
potential disasterinduced losses consequently housing recovery vital importance
overall restoration community relation recovery models help devising datadriven
policies better identify predisaster mitigation needs postdisaster recovery
priorities predicting possible outcomes different plans although several
recovery models proposed still gaps understanding decisions made individuals
different entities interact output recovery additionally integrating spatial
aspects recovery missing key many models current research proposes spatial model
simulation prediction homeowners recovery decisions incorporating recovery
drivers could capture interactions individual communal organizational decisions
recovus spatial agentbased model input data obtained publicly available data
sources model presented using data recovery staten island new york hurricane
sandy 2012 results confirm combination internal interactive external drivers
recovery affect households decisions shape progress recovery disaster recovery
recovery modeling agentbased modeling perceived community',
       'agentbased models equationbased models used model spread infectious
disease equationbased models shown capture overall dynamics disease outbreak
agentbased models able capture heterogeneous characteristics agents drive spread
outbreak however agentbased models computationally intensive capture advantages
equationbased agentbased models create hybrid model disease component hybrid
model switches agentbased equationbased switch determined using number agents
infected first test model town level county level investigating different switch
values geographic levels switching find hybrid model able save time compared
fully agentbased model without losing significant amount fidelity hybrid
agentbased equation based infectious disease simulation epidemiology'],
      dtype='<U1877')
```

3 Extract TF-IDF Features

[21]:
```
from sklearn.feature_extraction.text import TfidfVectorizer
```

[25]:
```
tf = TfidfVectorizer(ngram_range=(1, 2), min_df=2)
tfidf_matrix = tf.fit_transform(norm_corpus)
tfidf_matrix.shape
```

[25]:
```
(49, 1044)
```

[26]:
```
tfidf_matrix
```

[26]:
```
<49x1044 sparse matrix of type '<class 'numpy.float64'>'
        with 3658 stored elements in Compressed Sparse Row format>
```

4 Compute Pairwise Document Similarity

[27]:
```
from sklearn.metrics.pairwise import cosine_similarity
```

[28]:
```
doc_sim = cosine_similarity(tfidf_matrix)
doc_sim.shape
```

[28]:
```
(49, 49)
```

[29]:
```
doc_sim_df = pd.DataFrame(doc_sim)
doc_sim_df.head()
```

[29]:

	0	1	2	3	4	5	6	\
0	1.000000	0.094273	0.025776	0.014993	0.064654	0.093295	0.041820	
1	0.094273	1.000000	0.081363	0.006503	0.072477	0.107547	0.040347	
2	0.025776	0.081363	1.000000	0.097501	0.161255	0.174317	0.074181	
3	0.014993	0.006503	0.097501	1.000000	0.073408	0.106214	0.087259	
4	0.064654	0.072477	0.161255	0.073408	1.000000	0.083825	0.038487	

	7	8	9	…	39	40	41	42	\
0	0.043845	0.064278	0.040490	…	0.043683	0.042337	0.080045	0.202426	
1	0.031473	0.045665	0.138427	…	0.143750	0.313665	0.043438	0.061270	
2	0.030301	0.037159	0.076727	…	0.056104	0.050485	0.066391	0.016715	
3	0.023125	0.006142	0.059924	…	0.030495	0.015687	0.111510	0.050755	
4	0.048154	0.018480	0.062335	…	0.080851	0.050423	0.039466	0.036143	

	43	44	45	46	47	48
0	0.158292	0.077627	0.108755	0.062924	0.052884	0.084337
1	0.076427	0.009959	0.064723	0.038261	0.128178	0.036108
2	0.037922	0.030873	0.136950	0.081819	0.142849	0.058614
3	0.016923	0.019545	0.053004	0.046314	0.089196	0.068255
4	0.060866	0.097226	0.080929	0.114029	0.093980	0.039134

[5 rows x 49 columns]

5 Get list of Movie Titles

[30]:
```
papers_list = df['Title'].values
papers_list
```

[30]:
```
array(['Catch Me if You Can: Using a Threshold Model to Simulate Support for
Presidential Candidates in the Invisible Primary',
       'How Policy Decisions Affect Refugee Journeys in South Sudan: A Study
Using Automated Ensemble Simulations',
       'Methodological Issues of Spatial Agent-Based Models',
       'LevelSpace: A NetLogo Extension for Multi-Level Agent-Based Modeling',
       'Cascading Impacts of Payments for Ecosystem Services in Complex Human-
Environment Systems',
       '‘One Size Does Not Fit All’: A Roadmap of Purpose-Driven Mixed-Method
Pathways for Sensitivity Analysis of Agent-Based Models',
       'Calibrating Agent-Based Models with Linear Regressions',
       'Do Either Farm Characteristics or Social Dynamics Explain the Conversion
to Organic Farming by Dairy Farmers? An Agent-Based Model of Dairy Farming in 27
French Cantons',
       'An Agent-Based Model of Firm Size Distribution and Collaborative
Innovation',
       'An Innovative Approach to Multi-Method Integrated Assessment Modelling
of Global Climate Change',
       'Land-Use Changes in Distant Places: Implementation of a Telecoupled
Agent-Based Model',
       'Theory Development Via Replicated Simulations and the Added Value of
Standards',
       'Editorial: Meeting Grand Challenges in Agent-Based Models',
       'A Dynamic Computational Model of Social Stigma',
       'Estimating Spatio-Temporal Risks from Volcanic Eruptions Using an Agent-
Based Model',
       'Phase Transition in the Social Impact Model of Opinion Formation in
Scale-Free Networks: The Social Power Effect',
       'Do Farm Characteristics or Social Dynamics Explain the Conversion to
Organic Farming by Dairy Farmers? An Agent-Based Model of Dairy Farming in 27
French Cantons',
       'Tension Between Stability and Representativeness in a Democratic
Setting',
       'Homophily as a Process Generating Social Networks: Insights from Social
Distance Attachment Model',
       'The ODD Protocol for Describing Agent-Based and Other Simulation Models:
A Second Update to Improve Clarity, Replication, and Structural Realism',
       'Emergence of Small-World Networks in an Overlapping-Generations Model of
Social Dynamics, Trust and Economic Performance',
       'Metamodels for Evaluating, Calibrating and Applying Agent-Based Models:
A Review',
       'Computational Models That Matter During a Global Pandemic Outbreak: A
Call to Action',
       'A Software Architecture for Mechanism-Based Social Systems Modelling in
Agent-Based Simulation Models',
       'Impacts of Consensus Protocols and Trade Network Topologies on
Blockchain System Performance',
```

```
       'Simulating Crowds in Real Time with Agent-Based Modelling and a Particle
Filter',
       'Problem Solving: When Groups Perform Better Than Teammates',
       'A Weighted Balance Model of Opinion Hyperpolarization',
       "Comparing Actual and Simulated HFT Traders' Behavior for Agent Design",
       'An Agent-Based Approach to Integrated Assessment Modelling of Climate
Change',
       'Grade Language Heterogeneity in Simulation Models of Peer Review',
       'Reflexivity in a Diffusion of Innovations Model',
       'An Agent-Based Model for Simulating Inter-Settlement Trade in Past
Societies',
       'Price Formation in Parallel Trading Systems: Evidence from the Fine Wine
Market',
       'A Simulation Model of the Radicalisation Process Based on the IVEE
Theoretical Framework',
       'Model Exploration of an Information-Based Healthcare Intervention Using
Parallelization and Active Learning',
       'Indirect Reciprocity with Contagious Reputation in Large-Scale Small-
World Networks',
       'Modeling Cultural Dissemination and Divergence Between Rural and Urban
Regions',
       'WorkSim: An Agent-Based Model of Labor Markets',
       'Housing Market Agent-Based Simulation with Loan-To-Value and Debt-To-
Income',
       'Agent-Based Modelling of Values: The Case of Value Sensitive Design for
Refugee Logistics',
       'Leveraging Modularity During Replication of High-Fidelity Models:
Lessons from Replicating an Agent-Based Model for HIV Prevention',
       'The Unknown of the Pandemic: An Agent-Based Model of Final Phase Risks',
       'Seed Selection Strategies for Information Diffusion in Social Networks:
An Agent-Based Model Applied to Rural Zambia',
       'Halting SARS-CoV-2 by Targeting High-Contact Individuals',
       'Gregarious Behavior, Human Colonization and Social Differentiation: An
Agent-Based Model',
       'BEN: An Architecture for the Behavior of Social Agents',
       'RecovUS: An Agent-Based Model of Post-Disaster Household Recovery',
       'A Hybrid Agent-Based and Equation Based Model for the Spread of
Infectious Diseases'],
      dtype=object)
```

[31]:
```
papers_list.shape
```

[31]:
```
(49,)
```

6 Find Top Similar Papers for a Sample Paper

6.1 Find a sample paper

[34]:
```
paper_idx = np.where(papers_list == "Methodological Issues of Spatial␣
 ↪Agent-Based Models")[0][0]
paper_idx
```

[34]:
```
2
```

6.2 Get paper similarities

[35]:
```
paper_similarities = doc_sim_df.iloc[paper_idx].values
paper_similarities
```

[35]:
```
array([0.02577606, 0.08136321, 1.        , 0.09750124, 0.16125523,
       0.17431672, 0.07418075, 0.0303012 , 0.03715901, 0.07672667,
       0.08423844, 0.08116874, 0.29101922, 0.00649291, 0.07777895,
       0.02474062, 0.02948215, 0.03249036, 0.05889925, 0.08033719,
       0.01290792, 0.11114197, 0.06350669, 0.03839319, 0.03107078,
       0.10324467, 0.02235783, 0.07406834, 0.01430636, 0.03579566,
       0.06271778, 0.04000267, 0.14848624, 0.08022143, 0.12064449,
       0.08512844, 0.06020206, 0.0337683 , 0.05401322, 0.0561045 ,
       0.05048548, 0.06639055, 0.01671515, 0.03792188, 0.03087288,
       0.1369501 , 0.08181894, 0.142849  , 0.05861413])
```

6.3 Get top 5 similar paper IDs

[36]:
```
similar_paper_idxs = np.argsort(-paper_similarities)[1:6]
similar_paper_idxs
```

[36]:
```
array([12,  5,  4, 32, 47])
```

6.4 Get Top 5 similar papers

[38]:
```
similar_papers = papers_list[similar_paper_idxs]
print("Methodological Issues of Spatial Agent-Based Models")
print(similar_papers)
```

```
Methodological Issues of Spatial Agent-Based Models
['Editorial: Meeting Grand Challenges in Agent-Based Models'
 ''One Size Does Not Fit All': A Roadmap of Purpose-Driven Mixed-Method Pathways
for Sensitivity Analysis of Agent-Based Models'
 'Cascading Impacts of Payments for Ecosystem Services in Complex Human-
Environment Systems'
 'An Agent-Based Model for Simulating Inter-Settlement Trade in Past Societies'
 'RecovUS: An Agent-Based Model of Post-Disaster Household Recovery']
```

7 Build a paper recommender function to recommend top 5 similar papers for each paper

[40]:
```
def paper_recommender(paper_title, papers=papers_list, doc_sims=doc_sim_df):
    # find paper id
    paper_idx = np.where(papers == paper_title)[0][0]
    # get paper similarities
    paper_similarities = doc_sims.iloc[paper_idx].values
    # sort the similarities and get top 5
    similar_paper_idxs = np.argsort(-paper_similarities)[1:6]
    similar_papers = papers[similar_paper_idxs]
    # return the top 5
    return similar_papers
```

7.1 Get Some Paper Recommendations

[41]:
```
some_papers = papers_list[:10]
some_papers
```

[41]:
```
array(['Catch Me if You Can: Using a Threshold Model to Simulate Support for
Presidential Candidates in the Invisible Primary',
       'How Policy Decisions Affect Refugee Journeys in South Sudan: A Study
Using Automated Ensemble Simulations',
       'Methodological Issues of Spatial Agent-Based Models',
       'LevelSpace: A NetLogo Extension for Multi-Level Agent-Based Modeling',
       'Cascading Impacts of Payments for Ecosystem Services in Complex Human-
Environment Systems',
       '‘One Size Does Not Fit All’: A Roadmap of Purpose-Driven Mixed-Method
Pathways for Sensitivity Analysis of Agent-Based Models',
       'Calibrating Agent-Based Models with Linear Regressions',
       'Do Either Farm Characteristics or Social Dynamics Explain the Conversion
to Organic Farming by Dairy Farmers? An Agent-Based Model of Dairy Farming in 27
French Cantons',
       'An Agent-Based Model of Firm Size Distribution and Collaborative
Innovation',
       'An Innovative Approach to Multi-Method Integrated Assessment Modelling
of Global Climate Change'],
      dtype=object)
```

[43]:
```
for paper in some_papers:
    print("Paper:", paper)
    print("Top 5 Recommended Papers:", paper_recommender(paper_title=paper))
    print('-'*60)
```

```
Paper: Catch Me if You Can: Using a Threshold Model to Simulate Support for
Presidential Candidates in the Invisible Primary
Top 5 Recommended Papers: ['Emergence of Small-World Networks in an Overlapping-
```

```
Generations Model of Social Dynamics, Trust and Economic Performance'
 'Reflexivity in a Diffusion of Innovations Model'
 'The Unknown of the Pandemic: An Agent-Based Model of Final Phase Risks'
 'Homophily as a Process Generating Social Networks: Insights from Social
Distance Attachment Model'
 'An Agent-Based Model for Simulating Inter-Settlement Trade in Past Societies']
------------------------------------------------------------
Paper: How Policy Decisions Affect Refugee Journeys in South Sudan: A Study
Using Automated Ensemble Simulations
Top 5 Recommended Papers: ['Agent-Based Modelling of Values: The Case of Value
Sensitive Design for Refugee Logistics'
 'Simulating Crowds in Real Time with Agent-Based Modelling and a Particle
Filter'
 'Housing Market Agent-Based Simulation with Loan-To-Value and Debt-To-Income'
 'An Innovative Approach to Multi-Method Integrated Assessment Modelling of
Global Climate Change'
 'RecovUS: An Agent-Based Model of Post-Disaster Household Recovery']
------------------------------------------------------------
Paper: Methodological Issues of Spatial Agent-Based Models
Top 5 Recommended Papers: ['Editorial: Meeting Grand Challenges in Agent-Based
Models'
 '‘One Size Does Not Fit All’: A Roadmap of Purpose-Driven Mixed-Method Pathways
for Sensitivity Analysis of Agent-Based Models'
 'Cascading Impacts of Payments for Ecosystem Services in Complex Human-
Environment Systems'
 'An Agent-Based Model for Simulating Inter-Settlement Trade in Past Societies'
 'RecovUS: An Agent-Based Model of Post-Disaster Household Recovery']
------------------------------------------------------------
Paper: LevelSpace: A NetLogo Extension for Multi-Level Agent-Based Modeling
Top 5 Recommended Papers: ['Grade Language Heterogeneity in Simulation Models of
Peer Review'
 'Metamodels for Evaluating, Calibrating and Applying Agent-Based Models: A
Review'
 'Editorial: Meeting Grand Challenges in Agent-Based Models'
 'Leveraging Modularity During Replication of High-Fidelity Models: Lessons from
Replicating an Agent-Based Model for HIV Prevention'
 '‘One Size Does Not Fit All’: A Roadmap of Purpose-Driven Mixed-Method Pathways
for Sensitivity Analysis of Agent-Based Models']
------------------------------------------------------------
Paper: Cascading Impacts of Payments for Ecosystem Services in Complex Human-
Environment Systems
Top 5 Recommended Papers: ['Land-Use Changes in Distant Places: Implementation
of a Telecoupled Agent-Based Model'
 'Methodological Issues of Spatial Agent-Based Models'
 'BEN: An Architecture for the Behavior of Social Agents'
 'Editorial: Meeting Grand Challenges in Agent-Based Models'
 'Simulating Crowds in Real Time with Agent-Based Modelling and a Particle
Filter']
```

```
------------------------------------------------------------
Paper: ‘One Size Does Not Fit All’: A Roadmap of Purpose-Driven Mixed-Method
Pathways for Sensitivity Analysis of Agent-Based Models
Top 5 Recommended Papers: ['Editorial: Meeting Grand Challenges in Agent-Based
Models'
 'Metamodels for Evaluating, Calibrating and Applying Agent-Based Models: A
Review'
 'Methodological Issues of Spatial Agent-Based Models'
 'Land-Use Changes in Distant Places: Implementation of a Telecoupled Agent-
Based Model'
 'RecovUS: An Agent-Based Model of Post-Disaster Household Recovery']
------------------------------------------------------------
Paper: Calibrating Agent-Based Models with Linear Regressions
Top 5 Recommended Papers: ['A Hybrid Agent-Based and Equation Based Model for
the Spread of Infectious Diseases'
 'Metamodels for Evaluating, Calibrating and Applying Agent-Based Models: A
Review'
 'The ODD Protocol for Describing Agent-Based and Other Simulation Models: A
Second Update to Improve Clarity, Replication, and Structural Realism'
 '‘One Size Does Not Fit All’: A Roadmap of Purpose-Driven Mixed-Method Pathways
for Sensitivity Analysis of Agent-Based Models'
 'Leveraging Modularity During Replication of High-Fidelity Models: Lessons from
Replicating an Agent-Based Model for HIV Prevention']
------------------------------------------------------------
Paper: Do Either Farm Characteristics or Social Dynamics Explain the Conversion
to Organic Farming by Dairy Farmers? An Agent-Based Model of Dairy Farming in 27
French Cantons
Top 5 Recommended Papers: ['Do Farm Characteristics or Social Dynamics Explain
the Conversion to Organic Farming by Dairy Farmers? An Agent-Based Model of
Dairy Farming in 27 French Cantons'
 'Gregarious Behavior, Human Colonization and Social Differentiation: An Agent-
Based Model'
 'Reflexivity in a Diffusion of Innovations Model'
 'RecovUS: An Agent-Based Model of Post-Disaster Household Recovery'
 'Indirect Reciprocity with Contagious Reputation in Large-Scale Small-World
Networks']
------------------------------------------------------------
Paper: An Agent-Based Model of Firm Size Distribution and Collaborative
Innovation
Top 5 Recommended Papers: ['WorkSim: An Agent-Based Model of Labor Markets'
 'Computational Models That Matter During a Global Pandemic Outbreak: A Call to
Action'
 'Seed Selection Strategies for Information Diffusion in Social Networks: An
Agent-Based Model Applied to Rural Zambia'
 'Simulating Crowds in Real Time with Agent-Based Modelling and a Particle
Filter'
 'An Agent-Based Model for Simulating Inter-Settlement Trade in Past Societies']
------------------------------------------------------------
```

```
Paper: An Innovative Approach to Multi-Method Integrated Assessment Modelling of
Global Climate Change
Top 5 Recommended Papers: ['An Agent-Based Approach to Integrated Assessment
Modelling of Climate Change'
 'Simulating Crowds in Real Time with Agent-Based Modelling and a Particle
Filter'
 'A Software Architecture for Mechanism-Based Social Systems Modelling in Agent-
Based Simulation Models'
 'How Policy Decisions Affect Refugee Journeys in South Sudan: A Study Using
Automated Ensemble Simulations'
 'Land-Use Changes in Distant Places: Implementation of a Telecoupled Agent-
Based Model']
------------------------------------------------------------
```