

**VILNIAUS UNIVERSITETAS
FIZIKOS FAKULTETAS
TEORINĖS FIZIKOS IR ASTRONOMIJOS INSTITUTAS**

Justas Kvedaravičius

**ERDVINIŲ SKIRSTINIŲ POKYČIAI KINTANT EMPIRINIŲ DUOMENŲ
MASTELIUI**

BAKALAURO STUDIJŲ BAIGIAMASIS DARBAS

Fizikos studijų programa

Studentas	Justas Kvedaravičius
Leista ginti	2022-05-24
Darbo vadovas	dr. Aleksejus Kononovičius
Recenzentas	prof. dr. Egidijus Anisimovas
Instituto atstovas	prof. dr. Egidijus Anisimovas

Vilnius 2022

Turiny

1	Ivadas	2
2	Literatūros apžvalga	3
2.1	Ivairovės matavimo dydžiai	3
2.1.1	Standartinis nuokrypis	3
2.1.2	Gini koeficientas	4
2.1.3	Theil indeksas.....	5
2.2	Sukininiai modeliai	6
2.2.1	Rinkėjo modelis.....	7
2.2.2	Izingo modelis.....	8
3	Tyrimo metodika	13
3.1	Atsitiktinis modelis	13
3.2	Ivairovės indeksas	16
4	Tyrimo rezultatai	19
4.1	Ivairovės kitimas empiriniams duomenų masteliams	19
4.2	Ivairovės indekso verčių kitimas rinkėjo modeliui su triukšmu	22
4.3	Ivairovės indekso verčių kitimas Izingo modeliui pagal Metropolio interpretaciją..	24
4.4	Ivairovės indekso verčių kitimas Izingo modeliui pagal Kawasaki interpretaciją ...	27
5	Išvados	30
	Santrauka	31
	Summary	32
	Literatūros sąrašas	33

1 Įvadas

Fizikų bendruomenei pasinėrus į sudėtingų sistemų tyrinėjimą neilgai trukus buvo pradėti tirti socialiniai reiškiniai. Rinkėjų nuomonių dinamika, transporto srautai, getų formavimasis ar kalbos dialektų plitimas yra tik keletas sudėtingo socialinio pasaulio pavyzdžių. Naujai iškilusi tarpdisciplininių tyrimų sritis pavadinimu sociofizika ėmėsi tyrinėti tokius reiškinius statistinės fizikos metodais. Nors statistinės fizikos metodais įprasta tyrinėti atomus ar molekules, sociofizikoje modeliuojami žmonių ir jų grupių veiksmai [1]. Visgi pagrindinis socialinių sistemų fizikus dominantis klausimas išlieka toks pat - kaip dalelių tarpusavio sąveika iš atsitiktinių sistemos konfigūracijų sukuria tvarką ir dėsningumus [2].

Šiuo požiūriu nuomonių dinamika yra vienas plačiausiai tirtų reiškinių sociofizikoje. Ją tyrinėjantys mokslininkai siekia apibūdinti populiacijos būsenas nuomonės požiūriu bei kaip vyksta apsikeitimas jomis [2]. Tam buvo pasiūlyti įvairūs modeliai, iš kurių žinomiausias yra rinkėjo modelis. Jis panašus į statistinėje fizikoje plačiai ištirtą ir sąlyginai paprastą Izingo modelį. Taigi rinkėjo modelio populiarumą lėmė jo paprastumas ir patikrintas lankstumas teorinei analizei.

Vis dėlto [3] autoriai iškėlė klausimą - ar rinkėjo modelis yra tinkamas tyrinėti realių rinkėjų nuomonių dinamiką? Jie pastebi, jog teorinėms rinkėjo modelio išvadoms trūksta palyginimo su empiriniais ir eksperimentiniais duomenimis. Tokie duomenys pasižymi ne tik rinkėjų nuomonių kitimu laike, bet ir rinkėjų judėjimu erdvėje (geografijoje). Todėl [3] pasiūlo rinkėjo modelyje agentų sąveikas apibrėžti kaip vykstančias dvejose vietose - namuose arba darbe. Be to, erdvėje skirtingos nuomonės yra išsidėsčiusios nevienalytiškai. Remiantis šia prielaida [4] autorius žengia žingsniu toliau ir pasiūlo erdvinį rinkėjo modelį. Jame nuomonės išlieka pastovios, tačiau galimas rinkėjų erdvinis judėjimas.

Norint suprasti erdvės įtaką nuomonių dinamikai, reikia suprasti kaip duomenys gali parodyti rinkėjų ir jų nuomonių erdvinę struktūrą. **Šio darbo tikslas** - paaiškinti empirinių erdvinį skirstinių pokyčius kintant duomenų masteliui. Pagrindiniai darbo uždaviniai įgyvendinami šiame darbe:

1. Darbe bus apibendrinami empiriniai pastebėjimai padaryti profesinės praktikos metu ištyrus Jungtinės Karalystės 2011 m. gyventojų surašymo duomenis.
2. Analizė bus tęsiama su teoriniais sukininiais modeliais tiriant kaip erdviniai skirstiniai kinta keičiant mastelį juose.
3. Empirinių duomenų ir teorinių modeliavimų rezultatai bus palyginti tarpusavyje.

2 Literatūros apžvalga

2.1 Įvairovės matavimo dydžiai

Kasdienybėje naudojami žemėlapiai dažnai būna sudėti iš teritorijų - gatvių, kvartalų, miestų, regionų ar valstybių. Pasirinktos teritorijos savo mastelyje gali būti aprašomos kiekybiniais rodikliais - gyventojų tankiu, pragyvenimo lygiu. Fizikoje sutinkami erdviniai modeliai, pavyzdžiui, sukininiai, gali būti interpretuojami panašiai - kaip sudaryti iš teritorijų. Dvimatės gardelės ląstelėse galima patalpinti daleles, kurių būsenos būtų nusakomos kiekybiniu rodikliu - sukinio projekcija į statmeną gardelės plokštumai ašį. Žemose temperatūrose dalelių sukiniai suvienodėja ir susiformuoja to paties sukinio dalelių domenai. Tarsi žmonės, formuojantys kvartalus, miestus ar valstybes.

Tokių teritorijų tarpusavio įvairovę galime įvertinti pasinaudojant statistiniais rodikliais - imties pločiu, dispersija. Galima pasirinkti ir interpretavimo prasme sudėtingesnius dydžius, pavyzdžiui, entropiją. Socialinių mokslų disciplinose dažnai naudojami nelygybės matavimo dydžiai, Theil indeksas ar Gini koeficientas. Čia verta pastebėti, jog įvairovės matavimas konceptualiai panašus į nelygybės matavimą. Esant didesnei nelygybei, matuojamos vertės yra mažiau tarpusavyje panašios, taigi - įvairesnės.

2.1.1 Standartinis nuokrypis

Bene paprasčiausias ir dažnai sutinkamas duomenų sklaidos matavimo dydis yra standartinis nuokrypis. Jis apibrėžiamas iš kito statistinio dydžio žinomo kaip dispersija. Bendru atveju atsitiktiniams dydžiams, kurie yra pasiskirstę pagal normalųjį, beta ar kitokį žinomą skirstinį, galima apibrėžti jų statistinius momentus [5]:

$$\beta_k = \mathbb{E}[(X - \mathbb{E}[X])^k] \quad (1)$$

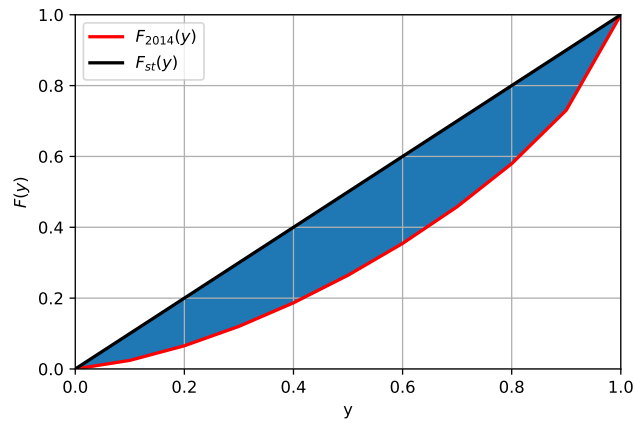
kur duomuo X žymi dominančio skirstinio atsitiktinį dydį. Taip apibrėžiami statistiniai momentai yra vadinami centriniais momentais ir β_2 žymi antrąjį momentą - dispersiją. Tokiu atveju standartinis nuokrypis yra apibrėžiamas taip:

$$\sigma = \sqrt{\beta_2}. \quad (2)$$

Vis tik įvertinant skirstinio statistines savybes iš empirinių duomenų rinkinio x_i naudojamosi patogesne išraiška:

$$\sigma = \sqrt{\frac{\sum_i^N (x_i - \mu)^2}{N - 1}}, \quad (3)$$

kur dydis μ žymi empirinių duomenų vidurkį. Tokiais atvejais išraiškoje yra būtinas nepaslinktas (angl. *unbiased*) daliklis $N - 1$, kuris žinomas kaip Beselio pataisa [6].



1 pav. Lorenz kreivės esant tolygiai pasiskirsčiusioms pajamoms $F_{st}(y)$ ir Lietuvos 2014 metų gyventojų pajamų pasiskirstymui $F_{2014}(y)$. Paveiklo plotas su mėlynu užpildu yra lygus pusei Gini koeficiento vertės (žr. (4) lygtį). $F(y)$ žymi sudėtinį pajamų pasiskirstymo tankį, o y - normalizuotą Lietuvos gyventojų skaičių.

2.1.2 Gini koeficientas

Vienas iš šiame darbe naudotų dydžių įvertinti pasiskirstymo įvairovei - Gini koeficientas. Taip pavadintas jo sukūrėjo italų statistiko ir sociologo Carrido Gini vardu. Šis koeficientas buvo pasiūlytas įvertinti populiacijos pajamų arba turto paskirstymo nelygybei, tačiau jo taikymas buvo išbandytas ir kartais tapdavo populiarius kitose disciplinose. Jo taikymą galima rasti vertinant išsilavinimo įgijamumą, sveikatos rodiklius, mokslo darbų citavimą [7]. Sociologinių tyrimų srityje [1] autoriai aptaria tyrimą, kuriame Gini koeficientu įvertinami internete prieinamų dainų atsisiuntimai priklausomai nuo aplinkos nuomonės poveikio.

Gini koeficientą galima apibrėžti pasinaudojant ekonomikoje sutinkamomis Lorenz kreivėmis. Kreivėms abscisių ašyje atvaizduojama sudėtinė populiacijos dalis y , o ordinačių - sudėtinei populiacijos daliai y tenkanti sudėtinė pajamų dalis $F(y)$ [8]. Jeigu visi individai nagrinėjamoje populiacijoje turėtų vienodas pajamas, tai būtų laikoma visiška pajamų pasiskirstymo lygybe, o $F(y)$ kreivė būtų tiesė. Taigi Gini koeficientas tokiu atveju yra plotų po visiškos lygybės ir empirinės Lorenz kreivės palyginimas. Nors originaliai C. Gini pasiūlė kitokį matematinį apibrėžimą šiam dydžiui, galima parodyti, jog Lorenz kreivių ir originalus apibrėžimai veda prie tos pačios analizinės išraiškos [9]. 1 paveiksle pateikiamas vizualus Lorenz kreivių ir Gini koeficiento atvaizdavimas - juoda linija žymi visiškos lygybės Lorenz kreivę, raudona žymi 2014 metų Lietuvos gyventojų pajamų pasiskirstymo Lorenz kreivę. Formali išraiška skaičiuoti koeficientui:

$$G = 2 \int_0^1 (y - F(y)) dy \quad (4)$$

čia y žymi sudėtinę populiacijos dalį, o $F(y)$ žymi sudėtinį pajamų tankį. Taigi Gini koeficientas yra 1 paveiksle mėlyna spalva pažymėto ploto santykis su visu po lygybės tiese esančiu plotu [10].

Iš 1 paveikslu galima pastebėti, jog su skirtingomis Lorenz kreivėmis galime gauti tą pačią indekso skaitinę vertę.

2.1.3 Theil indeksas

Kitas populiarus nelygybės matavimo vienetas žinomas Theil indekso pavadinimu. Kaip ir Gini koeficientas jis buvo pasiūlytas pajamų nelygybės matavimui, o pagrindinės jo idėjos kyla iš informacijos teorijos [11]. Savo forma jis yra panašus į Shannon informacijos entropijos išraišką. Theil indeksas tolydžių atsitiktinių dydžių atveju galėtų būti skaičiuojamas integruojant pagal duomenų skirstinio tikimybės tankio funkciją $p(x)$:

$$T = \int \frac{x}{\mu(p)} \ln \left(\frac{x}{\mu(p)} \right) dp(x), \quad (5)$$

čia μ žymi skirstinio vidurkį. Praktikoje dažniau susiduriama su diskrečiais duomenimis, o tokiu atveju patogiau naudotis šia išraiška:

$$T = \frac{1}{N} \sum_{i=1}^N \frac{x}{\mu} \ln \left(\frac{x}{\mu} \right). \quad (6)$$

Theil indeksas yra atskiras apibendrintos entropijos indekso atvejis [11]. Apibendrintas entropijos indeksas apibūdinamas tokia išraiška:

$$I_{GE}^{\alpha} = \frac{1}{\alpha(\alpha-1)} \int \left(\left(\frac{x}{\mu(p)} \right)^{\alpha} - 1 \right) dp(x), \quad (7)$$

čia α gali įgyti visas vertes bei lemia indekso jautrumą specifinėms dominančio pasiskirstymo sritimis, o $p(x)$ kaip ir 5 lygtyje reiškia tikimybės tankio skirstinio funkciją. Theil indekso atveju $\alpha = 2$.

Šiek tiek įžvalgų interpretuojant (6) formulę suteikia aksiomos iš informacijos teorijos [11]. Viena jų teigia, jog kažkokio įvykio p tikimybę galima interpretuoti kaip priešingą dydį informacijai apie įvykį. Kuo tikėtinėnis įvykis, tuo mažiau informacijos apie jį sužinoma. Šiuo atveju kuo santykinai mažiau x vertės tarpusavyje skiriasi, tuo jos artimesnės vidurkio μ vertei, o apibrėžtos tikimybės vertė būtų artimesnė vienetui. Informacijos funkcija (kuri yra logaritmas) duos artimą nuliui vertę. Tai galima interpretuoti taip: labiau tikėtinam įvykiui turima mažiau nežinomos informacijos apie jį. Lygiai taip pat [11] autoriai teigia, jog indeksu apskaičiuojama entropija yra tikėtina (nežinoma) skirstinio informacija.

Visgi vertinant (5) išraišką sudėtinga intuityviai nusakyti kaip vieno individo pajamos paveikia indekso vertę (Gini koeficientui tai pastebima [12] straipsnyje). Pagal [13] straipsnio autorius, tai žymiai aiškiau parodoma dispersijos skaičiavimo atveju. Individo pajamos keičia vidurkį ir dėl to gauname nenulinę dispersijos vertę. Theil indekso atveju [10, 13] straipsniu-

se teigiama, jog indekso išraiška leidžia išskaidyti jį ir suskaičiuoti pagal vertes visuomenės grupėms. Tai naudinga, nes gauti informaciją apie kiekvieno individo pajamas yra pernelyg sudėtinga ir kartais neetiška, tam pasitelkiamos įvairios duomenų rinkimo strategijos [7].

Kyla natūralus klausimas - kokios priežastys motyvuotų rinktis šiuos įvairovės dydžius nors statistiniai momentai gali suteikti panašią informaciją? Viena svarbi priežastis yra ta, jog (4) ir (5) formulėse pateikiami dydžiai yra netiesiniai ir jautresni vienai iš pajamų pasiskirstymo dalių [10]. Pavyzdžiui, Gini koeficientas yra jautriausias vidurinei populiacijos pajamų pasiskirstymo daliai. Apibendrintam entropijos indeksui tai keičiasi su skirtingomis α koeficiento vertėmis, kuo jos didesnės - tuo indekso vertė labiau kinta nuo pajamų, didesnių negu vidutinės. Tuo tarpu standartinis nuokrypis yra jautresnis išskirtims (angl. *outliers*).

2.2 Sukininiai modeliai

Įvairaus tipo socialinėse sistemose žmonės sąveikaudami tarpusavyje iš atsitiktinės būsenos gali sukurti tvarką. Socialinėse sistemose tai galima suprasti kaip politinio konsensuso arba poliarizacijos atsiradimą, kurį kiekybiškai galima įvertinti prie balsadėžių ar nuomonės apklausomis. Tokio pobūdžio tvarkos atsiradimas yra galimas dėl socialinės įtakos (angl. *social influence* arba *peer pressure*) reiškinių ir analogiškai fizikoje yra palyginamas su feromagnetiniais reiškiniais [2]. Taigi sukininiai modeliai yra puikūs kandidatai modeliuoti nuomonių dinamiką socialinėse sistemose.

Sukininiuose modeliuose įprasta sukinių sąveika gali pasirodyti pernelyg paprasta norint apibūdinti sudėtingus socialinius santykius ar pasirinkimus. Todėl pastebėdami šią problemą [1, 2] autoriai argumentuoja, kad sukininio modelio tinkamumą galima pritaikyti įvertinant sąveiką apibūdinančių kintamųjų verčių pasirinkimą. Viena vertus sukinių verčių kitimas gali būti ne šuoliškas, o tolygus, pats sąveikavimas ne lokalus, o globalus, arba kai kurių verčių kitimo gali visai nebūti. Nuomonių dinamikai modeliuoti yra pasiūlytos bent kelios iš esmės skirtingos sukininių sąveikų interpretacijos [14].

Daugelį sukininių modelių galima modeliuoti Monte Carlo metodu turint žinių apie jų būsenų tikimybių skirstinį. Pavyzdžiui, Izingo modelio atveju šiluminėje pusiausvyroje esančios sistemos būsenos tikimybė aprašoma pagal Bolcmano pasiskirstymą:

$$p_\mu = \frac{1}{Z} \exp\left(-\frac{E_\mu}{k_B T}\right), \quad (8)$$

čia Z - sistemos statistinė suma, k_B - Bolcmano konstanta, T - temperatūra, o E_μ žymi sistemos μ būsenoje energiją. Modeliuojamo virtualaus eksperimento rezultatai atitiks realios sistemos dinamiką šiluminėje pusiausvyroje (t.y. būsenų tikimybės pasiskirstymą pagal (8)) kai eksperimento sąlygose bus išpildytos ergodiškumo bei detalaus balanso sąlygos [15]. Ergodiškumas nurodo, kad bet kuri sistemos būsena turi būti pasiekama per baigtinį žingsnių skaičių. O

detalaus balanso sąlyga nustato, jog šiluminėje pusiausvyroje esančios sistemos tikimybės patekti į tam tikrą būseną yra lygios tikimybei ją palikti [15]. Tuomet sistemos perėjimo iš vienos būsenos į kitą intensyvumas ir tikimybės skirstinys turi būti susieti tokiu sąryšiu:

$$p_\mu P(\nu \rightarrow \mu) = p_\nu P(\mu \rightarrow \nu), \quad (9)$$

čia p_i žymi tikimybę sistemai būti atitinkamoje būsenoje, o $P(\nu \rightarrow \mu)$ žymi perėjimo intensyvumą iš ν būsenos į μ . Pritaikius šias sąlygas galima pasirinkti efektyvias šuolių perėjimo intensyvumo (angl. *acceptance ratios*) vertes [15].

2.2.1 Rinkėjo modelis

Rinkėjo modelis yra viena paprasčiausių sukininės sistemos veikimo interpretacijų. Iš tiesų jo pagrindinės taisyklės panašios į ląstelinio automato sistemos (angl. *cellular automaton*). Perėjimas iš vienos sistemos būsenos į kitą atliekamas taip: iš pradžių gardelėje pasirenkamas atsitiktinis sukinys ir jo artimiausias kaimynas. Tuomet pirmasis pasirinktas sukinys imituoja pasirinkto kaimyno būseną. Taigi pirmasis sukinys gali pakeisti savo būseną į priešingą arba likti toje pačioje. Bendru atveju įvykio, kurio metu būsena pakeičiama į priešingą, intensyvumas išreiškiamas taip:

$$P(s_k \rightarrow -s_k) = \frac{1}{2} \left(1 - \frac{s_k}{z} \sum_{\{j\}} s_j \right), \quad (10)$$

čia s_k žymi nagrinėjamos būsenos sukinio vertę, z žymi artimiausių kaimynų skaičių, o sumavimas $\{j\}$ atliekamas per artimiausius z kaimynų.

Sistemos būsenai apibūdinti galima įvesti porą makroskopinių tvarkos parametrų [1]. Vienas dydis yra skirtingas sukinio būsenas turinčių kaimynų skaičiaus santykis su visomis kaimynų poromis. Šis dydis savo prasme panašus į sistemos bendrą energiją. Kitas dydis, panašus į įmagnetėjimą, yra santykinių populiacijos dydžių turinčių skirtingus sukinius skirtumas. Modelyje atlikus vieną žingsnį magnetizacijos vertė pakinta, tačiau sukinų populiacijoms esant vienodo dydžio vidutiniškai per daugelį žingsnių ji nesikeičia [16].

Verta pastebėti, jog rinkėjo modelyje vienkrypčių sukinų domenų plėtimasis (angl. *coarsening*) yra skirtingas negu Izingo modelyje. Tai parodoma modeliuojant lašo formos vienu sukinų vertės regioną aplink esant priešingiems sukiniams (žr. 2 pav.) [1, 2]. Bėgant laikui regiono perimetras plinta, tačiau statistiškai lašo radiusas nesikeičia. Fizikine prasme tai reiškia, jog domeno plėtimasis kinta ne dėl paviršiaus įtempimo jėgos veikimo ir sistemos bandymo atsidurti mažiausios energijos būsenoje [2].

Visgi aptartas rinkėjo modelis neturi triukšmo savybės, pavyzdžiui, (10) lygtis nepriklauso nuo temperatūros ar kitokio parametro. Į modelį galima įvesti triukšmą, kuris bet kuriam pasirinktam sukiniui leidžia nepriklausomai nuo kaimynų pakeisti savo vertę. Vienas paprasčiausių būdų tai padaryti - prie (10) pridėti konstantą [16]. Triukšmas lemia tai, jog sistema

lieka į paramagnetinę panašioje būsenoje ir nesuvienodina visų savo sukinių [2]. Taip veikiantis modelis dažnai vadinamas rinkėjo modeliu su triukšmu (angl. *noisy voter model*). Remiantis detalaus balanso sąlyga (9) galima parodyti, jog tokiam rinkėjo modeliui tikimybių pasiskirstymas aprašomas ne Bolcmano skirstiniu parodytu (8) lygtyje, tačiau pagal Beta skirstinį [17]. Beta skirstinio parametrai α ir β priklauso nuo savaiminio nuomonės pasikeitimo (triukšmo) tikimybės. Kuo didesnė tikimybė savaime pakeisti sukinio vertę, tuo didesnė α vertė.

2.2.2 Izingo modelis

Prieš tai skyrelyje aptartas rinkėjo modelis buvo modelinė sistema, kurios elementarus veikimas gerai iliustruoja sukininių modelių veikimo principus. Toliau aptariamas Izingo modelis yra feromagnetinę sąveiką apibūdinantis modelis. Šios sąveikos Hamiltonianą galima užrašyti taip:

$$H = -J \sum_{\{i,j\}} s_i s_j, \quad (11)$$

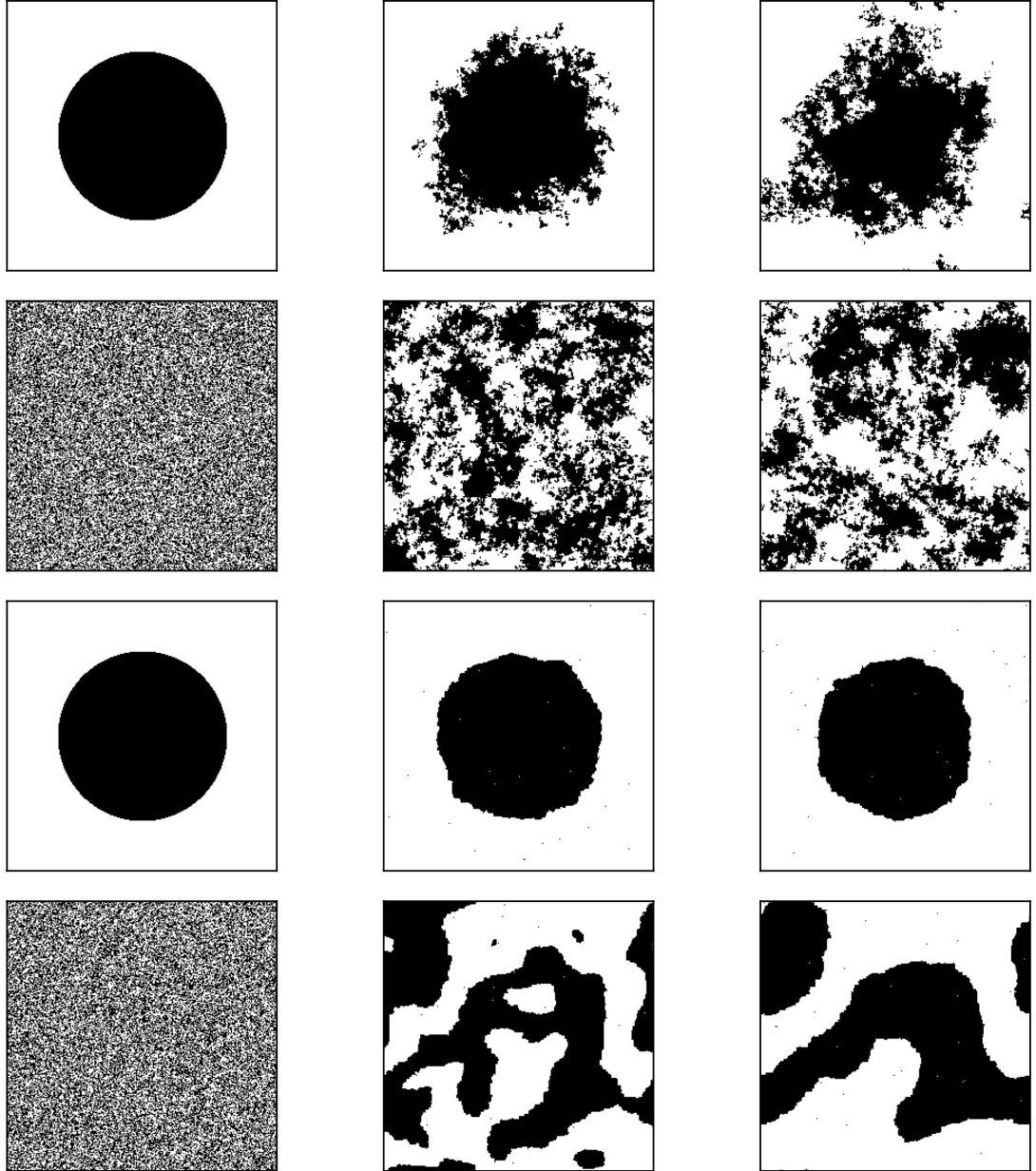
čia J žymi sukinių sąveikos energiją, $\{i, j\}$ žymi sumavimą per gardelėje esančius artimiausius kaimynus ir s_i, s_j žymi atitinkamas pasirinktų ir kaimyninių sukinių vertes. Jeigu sąveikos energija J yra teigiamas dydis, feromagnetinė sąveika skatina sukinius pasisukti ta pačia kryptimi [16].

Bendrai Izingo modelio veikimas yra panašus į aptartą rinkėjo modelyje, tačiau atsiranda esminiai skirtumai dėl to, jog šiuo atveju įtakos turi kaimynų dauguma. Sukinių pokyčiai gali būti įgyvendinami pagal Glauber arba Metropolio pasiūlytas dinamikas [15, 16, 18]. Metropolio pasiūlytoje dinamikoje iš pradžių pasirenkamas atsitiktinis sukinys. Tuomet yra įvertinamos jo sąveikos su kaimyniniais sukiniais. Jeigu sukinio apvertimas sumažintų bendrą sistemos energiją arba paliktų tokią pačią, toks pakeitimas priimamas. Kitais atvejais sukinio vertės pakeitimas sistemos energiją padidintų, todėl tai vyksta su tikimybe pagal Bolcmano pasiskirstymo (8) dėsnį. Šitai leidžia iš detalaus balanso sąlygos užrašyti būsenos priėmimo tikimybių santykį:

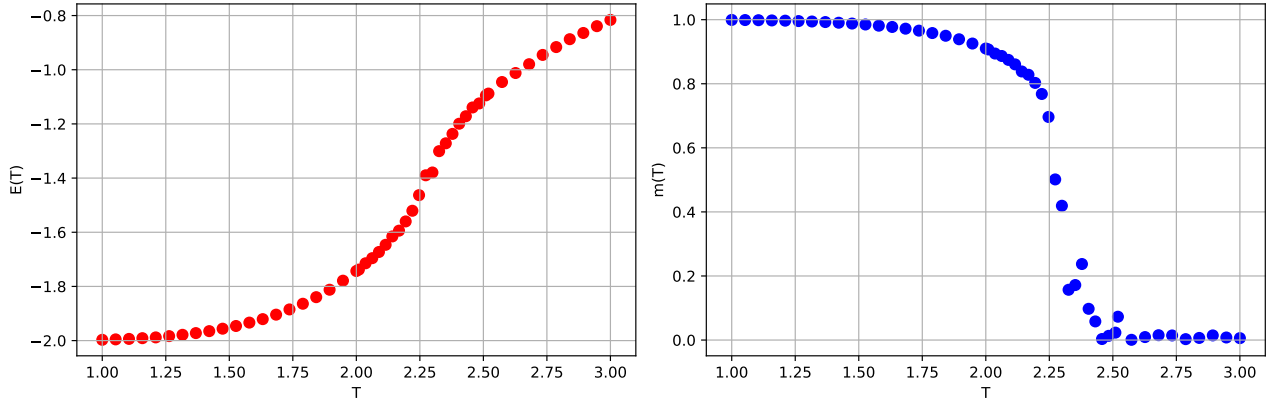
$$\frac{P(\nu \rightarrow \mu)}{P(\mu \rightarrow \nu)} = \frac{A(\nu \rightarrow \mu)}{A(\mu \rightarrow \nu)} = \exp(-\beta(E_\nu - E_\mu)), \quad \beta = \frac{1}{k_B T} \quad (12)$$

čia k_B yra Bolcmano konstanta, T žymi temperatūrą, o A žymi atitinkamų perėjimo tarp būsenų intensyvumo (angl. *acceptance ratios*) vertes. Modelis bus laikomas įgyvendinamu pagal Metropolio algoritmą, jeigu visi energijos mažėjimui palankūs sukinių apsisukimai bus priimami su tikimybe lygia 1, o sistemos energiją didinantys - pagal (12) lygtyje esančią eksponentės vertę [15].

Izingo modelio dinamika yra įdomi tuo, jog ji turi kritinį tašką T_C , feromagnetinėms medžiagoms žinomą kaip Kiuri temperatūrą. Temperatūrų vertėms mažesnėms negu šis taškas stebimas feromagnetinės sąveikos dominavimas ir vidutinė gardelės magnetizacija (tvarkos parametras) yra nenulinė, t.y. $\langle m \rangle > 0$. Tačiau atveju $T > T_C$ sistema yra paramagnetinėje



2 pav. Rinkėjo (1 ir 2 eilutės) ir Izingo (3 ir 4 eilutės) modelių konfigūracijos laiko t momentais 1, 256 ir 1024 (atitinkamai 1, 2 ir 3 stulpeliai). Gardelės dydis 256×256 , vienas laiko momentas lygus 256^2 įvykių, per kuriuos gali pasikeisti vieno sukinio vertė. Pirmoje ir trečioje eilutėje pavaizduoti rinkėjo modelio lašo išplitimas ir Izingo modelio lašo optimizacija link minimalios energijos. Antroje ir ketvirtoje eilutėje pavaizduota evoliucija nuo atsitiktinės konfigūracijos, rinkėjo modelio perimetras šiurkštus, o Izingo modelio perimetras darosi kuo tiesesnis (optimaliausioje energijos būsenoje). Izingo modelio temperatūra $T = T_0$, kur T_0 yra temperatūra normuota į Bolemano konstantą k_B ir feromagnetinės sąveikos konstantą J .



3 pav. Izingo modelio energijos E (kairėje) ir vidutinės magnetizacijos m (dešinėje) verčių priklausomybė nuo temperatūros T . Išreikšti į konstantas normuoti santykiniai vienetų dydžiai, kurie nusako modeliuojamos medžiagos charakteristikas. Izingo modeliui pagal Glauber interpretaciją kritinis taškas $T_C = 2.269T_0$.

būsenoje ir tvarkos parametras $m = 0$. Pačiame kritiniame taške T_C pasireiškia skalės invariantiškumas, kuomet žiūrint į sistemos gardelę skirtinga skyra (pakeičiant skalę) tvarkos parametrai savo vertės nekeičia [1]. O šio taško aplinkoje keičiant temperatūrą tvarkos parametrai kinta kaip laipsninės funkcijos, pavyzdžiui vidutinė magnetizacija:

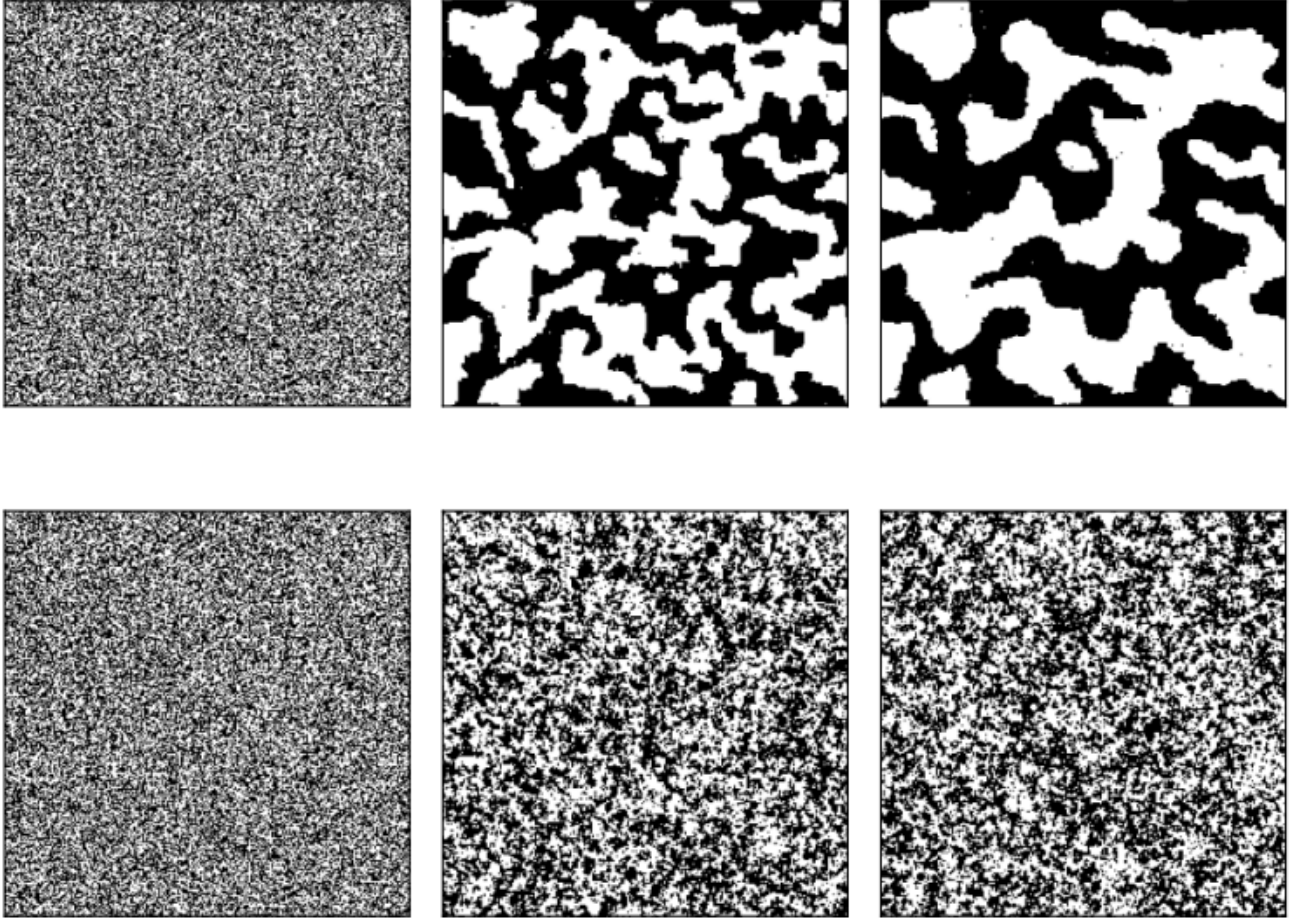
$$m = \frac{M}{N} \propto (T_C - T)^\gamma, \quad (13)$$

čia M žymi absoliučią magnetizaciją, N žymi sukinių kiekį, o γ yra kitimą apibūdinantis parametras. Šis kitimas matomas 3 paveiksle, tačiau jis egzistuoja ir kitiems termodinaminiais dydžiams su skirtingais laipsniniais parametrais [1]. Verta pastebėti, jog vidutinė sistemos magnetizacija m kiekybiškai neįvertina susidarančių domenų dydžio ar formų [2]. Siekiant tai įvertinti galima suskaičiuoti sistemos koreliacijos funkciją [2, 15, 16].

Analogiškai pagal rinkėjo modelio (10) lygtį galime užrašyti sukinio pakeitimo įvykio dažnius Izingo modelio interpretacijai pagal Glauber:

$$P(s_k \rightarrow -s_k) = \frac{1}{2} \left(1 - s_k \tanh \left(\frac{\beta J}{z} \sum_{\{j\}} s_j \right) \right), \quad (14)$$

čia simbolių žymėjimų prasmės tos pačios kaip (10) ir (12) lygtyse, $\{j\}$ žymi sumavimą per k sukinio kaimynus. Iš šios lygties galima padaryti svarbų pastebėjimą dėl skirtumo tarp rinkėjo ir Izingo modelio. Rinkėjo modelio atveju (10) lygtis gali būti tiksliai išsprendžiama bet kokiam dimensijų skaičiui, tačiau (14) atveju tai neįmanoma dėl kaimyninių sukinų priklausomybės pagal hiperbolinį tangentą [16]. Tačiau [1] pastebima, jog rinkėjo modelio tvarkinga būseną aukštesnėms ($d > 2$) dimensijoms pasiekama asimptotiškai ir tik baigtinėms gardelėms. Dar pagal [16] pastebimą (10) ir (14) lygčių (ne)tiesiškumo savybę galima atitinkamai pavadinti



4 pav. Izingo modelio pagal Kawasaki interpretaciją konfigūracijos laiko momentais $t = 1, 256$ ir 1024 (atitinkamai 1, 2 ir 3 stulpeliai). Pirmoje eilėje konfigūracijos, kai temperatūra $T = T_0$ ir $T = 3T_0$ antroje eilėje. Čia T_0 yra normuota į Bolcmano ir feromagnetinės sąveikos konstantas.

sukinio vertės pakeitimo taisyklės. Rinkėjo modeliu ji buvo proporcinė (priklausė nuo kaimynų vienkrypčių sukinių proporcijos), o Izingo - daugumos (kaimyninių sukinių dauguma nusprendžia koks žingsnis galimas). Kai kurie tyrėjai išskirtinai dažniau remiasi į daugumos balsavimą panašiomis taisyklėmis [19].

Iki šiol aprašyta Izingo modelio dinamika buvo paremta sukinių verčių keitimusi pagal sistemos energijos mažinimą. Tačiau įmanoma ir kitokia sukininio modelio interpretacija, kurioje sukiniai nekeičia savo verčių, o keičiasi vietomis. Viena iš vadinamos Kawasaki interpretacijos savybių - absoliučios skirtingų sukinių populiacijos nekinta, todėl magnetizacija prisijungia prie temperatūros kaip sistemos parametras (angl. *conserved-order parameter*) [15].

Šiai interpretacijai galima užrašyti įvykių intensyvumo lygtį analogiška prieš tai nagrinėtoms (10) ir (14) lygtims:

$$P(s_j, s_k) = \frac{1}{2} \left(1 - \tanh(2\beta J) \left(s_j \sum_{\{i\}, i \neq k} s_i + s_k \sum_{\{i\}, i \neq j} s_i \right) \right) \cdot \frac{1}{2} (1 - s_j s_k), \quad (15)$$

kur β yra parametras priklausantis nuo temperatūros, J yra feromagnetinės sąveikos energija, s_j ir s_k žymi kaimyninius sukinius, kurie yra sukeičiami vietomis, o sumavimas $\{i\}$ vyksta pagal visus kaimyninius sukinius neįskaitant tik vieno, su kuriuo būtų keičiamasi vietomis. Šioje lygtyje atsiranda papildomas daugiklis, kadangi pasirinkti sukiniai s_j ir s_k turi būti priešingi [16].

Kompiuterinės šio modelio simuliacijos vyksta gerokai lėčiau negu Izingo pagal Metropolisio algoritmą, kadangi kaimyninių dalelių keitimasis vietomis yra difuzinio pobūdžio judėjimas. Tai pastebdami [15] autoriai siūlo pakeisti sukinių apsikeitimo vietomis taisyklę leidžiant keistis bet kokiems skirtingiems sukiniams visoje gardelėje priimant energiją didinančius įvykius pagal (12) lygties tikimybę, o mažinančius ar nekeičiančius - su tikimybe lygia 1. Tokiu atveju dinamika sistemai judant link pusiausvyros būsenos skiriasi, tačiau pasiekiamas ta pati būsenas.

3 Tyrimo metodika

3.1 Atsitiktinis modelis

Empirinių erdvinių skirstinių pokyčių analizei pasirinkti šiuo metu naujausi (2011 metų) viešai prieinami Jungtinės Karalystės gyventojų surašymo duomenys¹. Pasinaudojant šiais empiriniais duomenimis galima iliustruoti erdvinės struktūros pavyzdį ir skirstinių pokyčius kintant duomenų masteliui.

Šiame surašyme iš viso prieinami 24 Jungtinės Karalystės geografinių vienetų (teritorijų) suskirstymo masteliai ir 27 populiacijos demografiniai rodikliai. Iš jų buvo pasirinkti 9 masteliai bei 6 demografiniai rodikliai. Plačiau šie duomenys pristatomi 5 paveikslo rango-dydzio kreivių grafikuose. Šios kreivės nuo sudėtinės tikimybės tankio funkcijos grafiko skiriasi ašių pasirinkimu - jos sukeistos vietomis. Abscisių ašyje esantis normalizuotas rangas atitinka tikimybę įgyti šiam rangui lygią arba didesnę ordinačių ašies vertę. Toks vaizdavimas sugalvotas atvaizduoti miestų dydžiams, o sociofizikoje pasitelktas rinkiminių duomenų analizei [20].

Rango-dydzio kreivių paveikslo legendoje yra pažymėti skirtingų mastelių (nuo smulkiausio iki stambiausio) pavadinimai. Galima pastebėti, jog stambėjant masteliui šių kreivių statusas mažėja. Vis dėlto šis mažėjimas nėra vienodas skirtingiems demografiniams rodikliams esantiems 5 paveiksle. Pavyzdžiui, lyties ar darbo klasės skirstinių kreivės nuo mastelio pokyčių kinta nežymiai, tačiau etniškumo ar kilmės šalies rodikliams tai matyti akivaizdžiau. Rango-dydzio kreivių formą skirtingiems masteliams galima palyginti įvertinant įvairovės dydžių vertes kiekvienam masteliui bei apskaičiuojant jų tarpusavio santykinį kitimą. Norint įvertinti šio santykinio kitimo prigimtį, suformuokime atsitiktinį modelį, su kuriuo tai bus palyginama.

Šį modelį sukonstruokime tokį, kad erdviniai (teritoriniai) vienetai būtų grupuojami ne su jų artimiausiais kaimynais, o atsitiktiniu būdu. Ši procedūra statistikoje žinoma kaip nulinės hipotezės (angl. *null-hypothesis*) tikrinimo procedūra. Todėl šis modelis toliau tekste bus vadinamas nuliniu arba atsitiktiniu modeliu. Modelio veikimą galime apibūdinti keliais žingsniais.

Pirma, iš turimų duomenų pasirenkamas smulkiausias įmanomas mastelis. Pavyzdžiui, empirinių duomenų atveju (žr. 1 lentelę) tai atitiktų 34753 erdvinių vienetų mastelį. Analogiškai sukininio modelio atveju smulkiausias mastelis būtų kvadratinės gardelės matrica, o erdvinių vienetų skaičius - L^2 , kur L - gardelės taškų skaičius vienoje kraštinėje.

Antra, šio mastelio erdviniai vienetai yra atsitiktiniu būdu sujungiami į naujas grupes, kurios sudaro stambesnius teritorinius vienetus, o viso sudarytų vienetų skaičius lygus stambesnio mastelio teritorinių vienetų skaičiui. Empirinių duomenų atveju - tai atitiktų 7201 teritorinių vienetų mastelį. Sukininių modelių atveju nauji erdviniai vienetai sujungiami iš ankstesnio detalesniojo mastelio 4 vienetų.

Trečia, tokiu būdu sudarytų mastelių erdviniams vienetais iš naujo suskaičiuojamos de-

¹Prieiga internetu: https://www.nomisweb.co.uk/sources/census_2011_ks

mografinių rodiklių vertės arba įvertinama magnetizacija. Šiame darbe empirinių duomenų rodiklių vertės buvo iš naujo įvertintos pagal absoliučius dydžius atsižvelgiant į Simpsono paradokso įtaką vidurkiams. O paklaidos empiriniams duomenims įvertintos savirankos (angl. *bootstrap*) metodu. Pagal naujai įvertintas vertes galima suskaičiuoti įvairovės dydžių vertes. Empiriniams duomenimis jos naudojamos keturios: imties plotis, standartinis nuokrypis, Gini koeficientas ir Theil indeksas.

1 lentelė. Empirinei analizei pasirinkti duomenų masteliai iš Jungtinės Karalystės 2011 m. gyventojų surašymo.

Pavadinimas anglų kalba	Trumpinys	Erdvinių vienetų kiekis (vnt.)	Vidutinis gyventojų skaičius erdviniame vienetė (vnt., suapvalinti)
2011 super output areas - lower layer	lsoa2011	34753	1613
2011 super output areas - mid layer	msoa2011	7201	7787
postcode districts	pd2011	2269	24713
parliamentary constituencies 2010	pca10	573	102959
local authorities: district / unitary	ualad09	348	163614
local authorities: county / unitary	uacounty09	174	327228
postcode areas	pt2011	105	534056
regions	gor	10	5607591
country	country	2	28037956

Erdvinių vienetų grupavimo procedūrai pagal atsitiktinį modelį galioja centrinė ribinė teorema. Jos galiojimas leidžia nuspėti kokių dėsningumu keisis naujai suformuotiems masteliams apskaičiuojamos kai kurių įvairovės dydžių vertės. Standartinio nuokrypio verčių kitimą skirtingiems masteliams galima parodyti analiziškai.

Tegul dydis K žymi į vieną erdvinį vienetą pagal atsitiktinį modelį sujungtų teritorijų skaičių. Tuomet vieno naujai suformuoto erdvinio vieneto demografinis rodiklis arba magnetizacija bus atsitiktinis dydis X . Jis savo ruožtu yra smulkesnių teritorinių vienetų dydžių x_i vidurkis:

$$X = \frac{1}{K} \sum_{i=1}^K x_i. \quad (16)$$

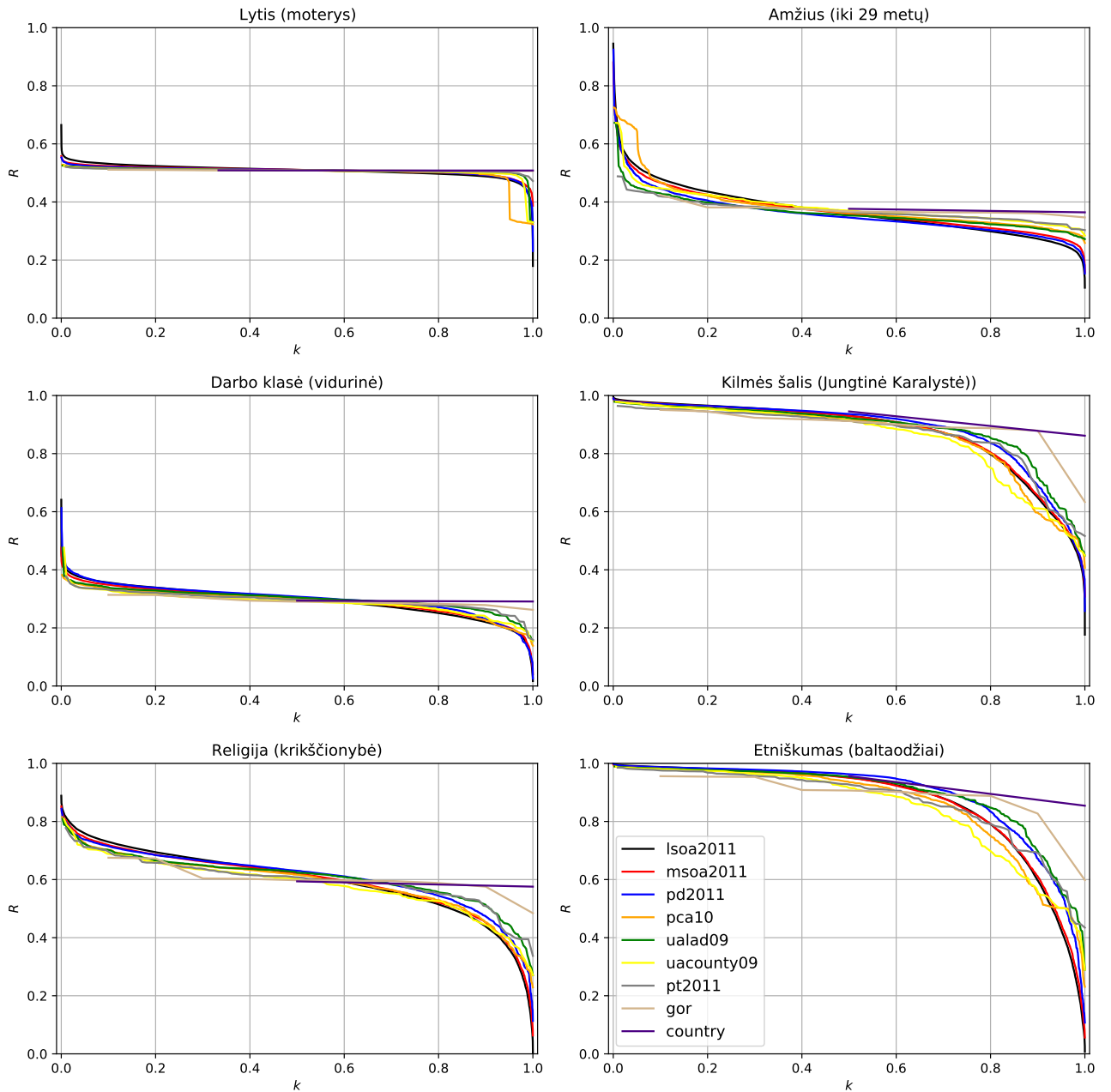
Galima apskaičiuoti X dispersiją:

$$\text{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}\left[\left(\frac{1}{K} \sum_{i=1}^K x_i - \mathbb{E}\left[\frac{1}{K} \sum_{i=1}^K x_i\right]\right)^2\right]. \quad (17)$$

Iš paskutiniojo nario gaunama x_i sumos dispersija ir iškeltas daliklis K^2 . Pritaikius centrinę ribinę teoremą:

$$\frac{1}{K^2} \text{Var}\left[\sum_{i=1}^K x_i\right] = \frac{1}{K^2} \left(\sum_{i=1}^K \text{Var}[x_i] + \sum_{i=1}^K \sum_{j \neq i}^K \text{Cov}[x_i, x_j]\right). \quad (18)$$

Jeigu dydžiai x_i yra nepriklausomi ir vienodai pasiskirstę, tuomet kovariacijos narys lygus nuliui.



5 pav. Pasirinktų empirinių duomenų demografinių rodiklių atvaizdavimas rango-dydžio kreivėse. R žymi demografiniai grupei priklausančių asmenų dalį teritoriniame vienetė, vertės yra iš intervalo $(0; 1)$. k žymi santykinį skirstinio rangą. Skirtingomis spalvomis pažymėtos kreivės atitinka skirtingus mastelius. Legendoje pateikti mastelių žymėjimų trumpiniai aptariami 1 lentelėje.

Taigi gaunama tokia dispersijos išraiška:

$$\text{Var}[X] = \frac{1}{K} \text{Var}[x_i], \quad (19)$$

ir atitinkamai skaičiuojant standartinį nuokrypį:

$$\sigma_X = \sqrt{\text{Var}[X]} = \sqrt{\frac{\text{Var}[x_i]}{K}} = \frac{\sigma_{x_i}}{\sqrt{K}}. \quad (20)$$

Gauta priklausomybė nusako pagal kokį dėsnį kinta atsitiktinio modelio būdu sudarytų mastelių standartinio nuokrypio (kaip įvairovės matmens) vertės. Kartu galima palyginti pagal empirinių duomenų mastelius gaunamą standartinio nuokrypio kitimą ir nustatyti ar stebimų struktūrų prigimtis yra atsitiktinė ar ne.

3.2 Įvairovės indeksas

Žinojimas, jog stebima dinamika yra atsitiktinio pobūdžio, gali suteikti kokybinę informaciją apie tvarkos (ne)buvimą nagrinėjamoje empirinėje ar sukininėje sistemoje. Tačiau remiantis tiktais modelio pagalba yra sudėtinga pasakyti kiek dinamika artima visiškai tvarkai ar netvarkai. Pavyzdžiui, nustatant įvairovės dydžių kitimą empiriniams masteliams nepavyktų nusakyti kaip tiksliai rodikliais pasižyminės populiacijos yra atskirtos į dvi skirtingas grupes. Tai analogiškai 2 paveiksle galima matyti sukininiams modeliams. Vien tik su atsitiktiniu modeliu negalima kiekybiškai palyginti rinkėjo ir Izingo modelių konfigūracijų įvairovės (tvarkingumo) po 1024 iteracijų (žr. 2 paveikslą trečią stulpelį).

Kiekybiškam įvairovės (arba tvarkos) reiškinio įvertinimui galima įsivesti įvairovės indeksą. Prieš detalesnį jo apibrėžimą reikia atkreipti dėmesį, jog įvairovę nusakančiais dydžiais bus vadinami 2.1 skyriuje paminėti ir apibrėžti dydžiai, o įvairovės indeksas bus apibrėžtas toliau. Taigi sudarykime jį tokiu būdu, kad jis galėtų nusakyti tris būsenas - netvarką, poliarizaciją ir tvarkingą išsimaišymą. Šių savokų apibrėžimas gali skirtis tarp įvairių disciplinų [2, 14], tačiau poliarizacija čia laikoma tuo atveju, kai skirtingos grupės yra visiškai atskirtos. Šios būsenos dvimatei gardelei pavaizduotos 6 paveiksle.

Įvairovės indekso skaičiavimą apibrėžiame per santykinį įvairovę nusakančių dydžių verčių kitimą (žr. 3.1 skyrelį). Kadangi jau žinome kaip standartinis nuokrypis kinta atsitiktinio išsimaišymo atveju, įvertinkime įvairovės indekso vertę naudodami šį dydį. Pirmiausia, smulkausio mastelio duomenims apskaičiuojama standartinio nuokrypio vertė. Visų stambesnių mastelių apskaičiuotos nuokrypio vertės yra sunormuojamos į smulkiausiąją. Atvaizdavę tokių normuotų verčių priklausomybę nuo mastelio parametrų gautume, jog vertės mažėja nuo 1 link 0 (žr. 7 paveikslą). Tuomet įvairovės indekso vertė bus skaičiuojama kaip šia kreive apribotas plotas, t.y. skaičiuojamas integralas. Labai panašiai skaičiuojamas ankstesniame skyriuje



6 pav. Įvairios indeksu vertės atitinkančių būsenų konfigūracijos sukininėse sistemose. Kairioji $I = -1$ atitinka tvarkingą išsimaišymą, vidurinioji $I = 0$ atitinka atsitiktinio modelio išsimaišymą, o dešinioji konfigūracija $I = 1$ atitinka poliarizaciją.

aptartas Gini koeficientas.

Tarkime gardelės konfigūracijoje turime atsitiktinio modelio atvejį, t.y. $\sigma(s) = 1/\sqrt{s}$. Šiuo atveju gauname indekso vertę lygią 0 (6 paveikslo vidurinis atvejis). Tuomet skaičiuojant integralą (plotą po atsitiktinio modelio $\sigma(s)$ priklausomybę) gaunama tokia funkcinė priklausomybė:

$$A_{base} = \int_1^{s_{max}} \sigma(s) ds = \int_1^{s_{max}} \frac{1}{\sqrt{s}} ds = 2(\sqrt{s_{max}} - 1). \quad (21)$$

čia $\sigma(s)$ žymi santykinio standartinio nuokrypio kreivę pagal 7 paveikslą, o s_{max} atitinka stambiausiame mastelyje į vieną erdvinį vienetą apjungtų sukinų skaičių. Čia s_{max} yra panašus į K dydžiui analogišką K_{max} (žr. prieš tai skyrelį). Norint atsitiktinio modelio atveju gauti indekso vertę lygią 0, reikia ją skaičiuoti taip:

$$I = A_{data} - A_{base}, \quad (22)$$

kur A_{data} žymi nagrinėjamų duomenų atvejį su unikaliu pagal duomenis santykinio standartinio nuokrypio vertės kitimu $\sigma(s)$:

$$A_{data} = \int_1^{s_{max}} \sigma(s) ds. \quad (23)$$

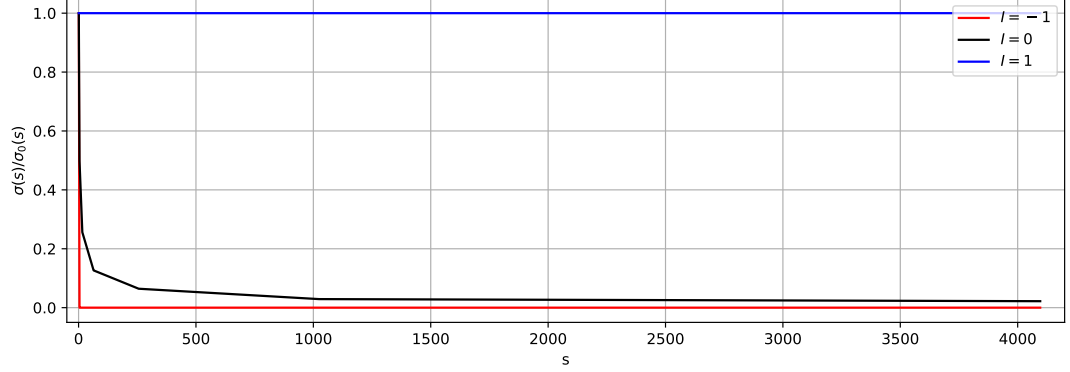
Toliau įvertinkime 6 paveikslo dešinįjį atvejį, kuomet sistemos konfigūracija atitinka būseną, kurioje skirtingi sukiniai yra visiškai atskirti į nepriklausomus domenų. Ši būsena be to atitiktų ir konsensuso atvejį, o jai norime turėti vertę lygią 1. Plotas po 7 paveikslo atitinkama kreive:

$$A_{max} = \int_1^{s_{max}} \sigma(s) ds = \int_1^{s_{max}} 1 ds = s_{max} - 1. \quad (24)$$

Tam, jog būtų gautas norimas dydis, reikalinga šią lygtį skaičiuoti tuo pačiu principu kaip ir

atsitiktinio modelio atveju bei sunormuoti:

$$I = \frac{A_{data} - A_{base}}{A_{max} - A_{base}}. \quad (25)$$



7 pav. Skirtingas įvairovės indekso I vertės atitinkančių santykinio standartinio nuokrypio kreivės. Dydis $\sigma(s)/\sigma_0(s)$ žymi standartinio nuokrypio santykinę vertę pagal smulkiausio mastelio vertę. Pavaizduotos priklausomybės gautos iš kvadratinės gardelės, kurios dimensijos - 128x128.

Priešingu, tvarkingo išsimaišymo, atveju norėtume gauti indekso vertę -1. Bandant pasiekti šį rezultatą svarbu atkreipti dėmesį į tai, kad plotai po kreivėmis nėra simetriški atsitiktinio modelio kreivės atžvilgiu. Tai matyti grafiškai 7 paveiksle. Kitaip tariant:

$$A_{max} - A_{base} \neq A_{base} - A_{min}. \quad (26)$$

A_{min} žymi integralo vertę tvarkingo išsimaišymo atveju:

$$A_{min} = \int_1^{s_{max}} \sigma(s) ds = \int_1^{s_{max}} \delta(1-s) ds. \quad (27)$$

Analiziškai pagal 7 paveiksle matoma kreivę šio integralo vertė bus lygi 0, tačiau realiu atveju gausime mažą teigiamą skaičių.

Taigi indekso vertės normavimas priklausys nuo to kaip turima kreivė A_{data} lyginasi su atsitiktinio modelio kreive A_{base} . Tuomet įvairovės indekso apskaičiavimo formulė:

$$I = \begin{cases} \frac{A_{data} - A_{base}}{A_{max} - A_{base}} & \text{jeigu } A_{data} > A_{base}, \\ \frac{A_{data} - A_{base}}{A_{base} - A_{min}} & \text{kitais atvejais.} \end{cases} \quad (28)$$

4 Tyrimo rezultatai

4.1 Įvairovės kitimas empiriniams duomenų masteliams

Pasinaudojant atsitiktinio modelio metodika, kuri yra aprašyta 3.1 skyriuje, galima nustatyti ar empiriniuose masteliuose užfiksuota demografiniais rodikliais pasižyminti populiacijos dalis yra pasiskirsčiusi atsitiktinai. Kadangi žinoma koku dėsniu turi kisti atsitiktinio modelio savybėmis pasižymintis duomenys, palyginimui būtina nustatyti santykinį įvairovę nusakančių dydžių kitimą empiriniams duomenims.

Jų kitimas yra pateiktas 8 paveiksle. Kiekvienam demografiniam rodikliui yra įvertinti 4 įvairovės dydžiai: imties plotis, standartinis nuokrypis, Gini koeficientas ir Theil indeksas. Galima pastebėti, jog skirtingiems rodikliams santykinis pokytis skiriasi keliomis eilėmis. Didžiausias santykinis kitimas yra pagal lytį ir darbo klasę, o mažiausias - pagal etniškumą ir kilmės šalį. Galima teigti, jog įvairovė stambesniuose masteliuose lyčiai ir darbo klasei sumažėja labiau negu kitiems demografiniams rodikliams.

Taip pat verta pastebėti, jog įvairovę nusakančius dydžius galima išskirti į dvi grupes pagal santykinio kitimo kreivės statumą. Pirmoji susideda iš standartinio nuokrypio ir Gini koeficiento, o antroji iš imties pločio ir Theil indekso. Šių grupių kraštiniai atvejai buvo pasirinkti tolimesnei analizei. Theil indeksas vietoje imties pločio, kadangi jis mažiau jautrus išskirtims, bei standartinis nuokrypis vietoje Gini koeficiento dėl jo paprastesnio suskaičiavimo būdo.

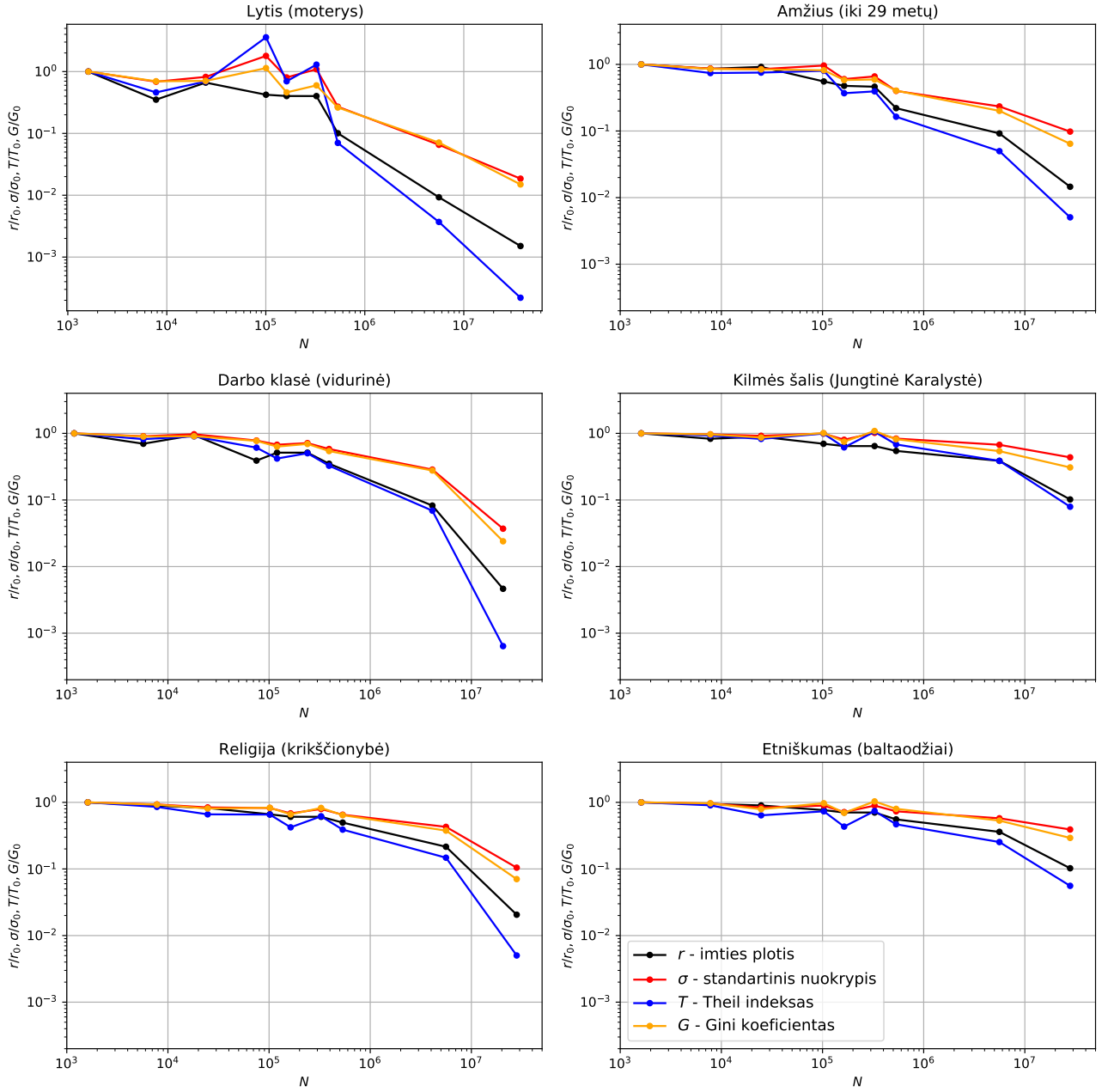
Lyties demografinio rodiklio 8 paveiksle esančiame grafike stebimas santykinis įvairovę nusakančio dydžio padidėjimas. Erdvinių vienetų grupavimo procedūrai toks atvejis yra ganėtinai neįprastas, nes jos būdu nėra įmanoma gauti įvairovės matavimo dydžio padidėjimo. Tai reiškia, jog ši anomalija yra susijusi su duomenų pobūdžiu. Iš tiesų 1 lentelėje pateikiami masteliai nėra griežtai sudaryti vienas iš kito, todėl nukrypimai yra įmanomi.

Vienas svarbesnių rezultatų yra empirinių duomenų įvairovės kitimo palyginimas su rezultatais gaunamais pagal atsitiktinį modelį. Vizualiai šie pokyčiai standartiniam nuokrypiui ir Theil indeksui yra palyginti atitinkamai 9 ir 10 paveiksluose.

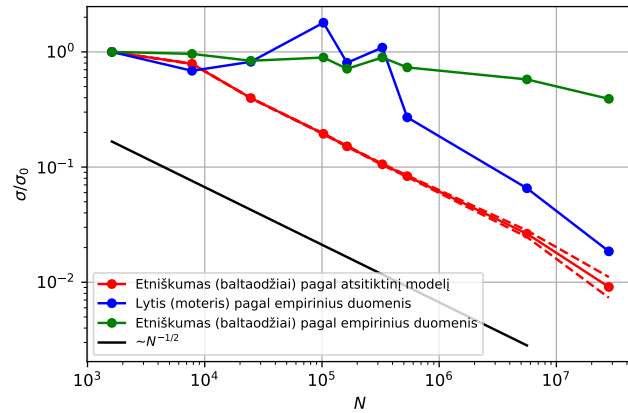
Standartinio nuokrypio kitimas atsitiktiniam modeliui buvo parodytas 3.1 skyriuje ir 9 paveiksle pavaizduotas raudona kreive. Juoda tiesė žymi patį proporcingumo dėsnį $K^{-1/2}$. Taip pat pasirinkti demografiniai rodikliai pagal lytį ir etniškumą yra didžiausio ir mažiausio santykinio standartinio nuokrypio kitimo pavyzdžiai.

Theil indekso kitimas pateiktas 10 paveiksle. Šiuo atveju atsitiktinio modelio rezultatai duoda kitokį dėsnį, tiksliau K^{-1} . Šis kitimas buvo nustatytas skaitmeniškai, tačiau kartu su darbo vadovu buvo aptartas ir analizinis šio rezultato išvedimas.

Lyginant atsitiktinio modelio ir empirinių duomenų kreives buvo pereita nuo erdvinių vienetų sugrupavimo dydžio K prie N , kuris žymi vidutinį erdviniam vienetui tenkantį gyventojų skaičių. Pagal 1 lentelės duomenis nesunku matyti, kad $K \approx N/N_0$, čia N_0 žymi smulkiausio

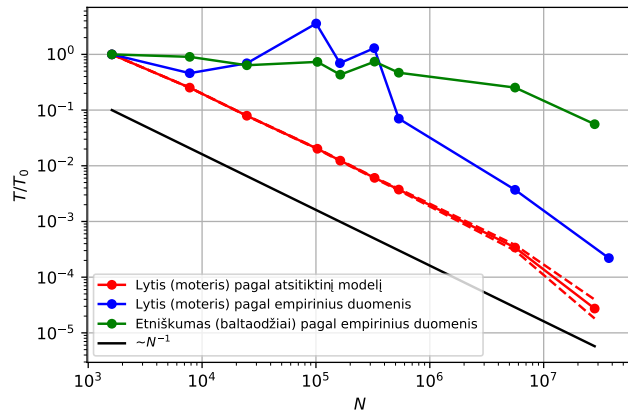


8 pav. Įvairovę aprašančių dydžių santykinis pokytis keičiantis empirinių duomenų masteliui. R žymi įvairovės matavimo dydžio vertės santykį su to paties dydžio detaliausio mastelio verte, N vidutinį gyventojų skaičių vienam erdviniam vienetui.



9 pav. Standartinio nuokrypio verčių kitimas keičiant duomenų mastelį. σ/σ_0 žymi skirtingų demografinių rodiklių normuotas standartinio nuokrypio vertės (lyginant su smulkiausio mastelio verte), N žymi mastelio vidutinį teritoriniam vnt. tenkantį gyventojų skaičių. Punktyrinėmis linijomis pažymėtas atsitiktinio modelio kreivės 95% patikimumo intervalas.

mastelio vidutinį gyventojų skaičių teritoriniam vienetui.

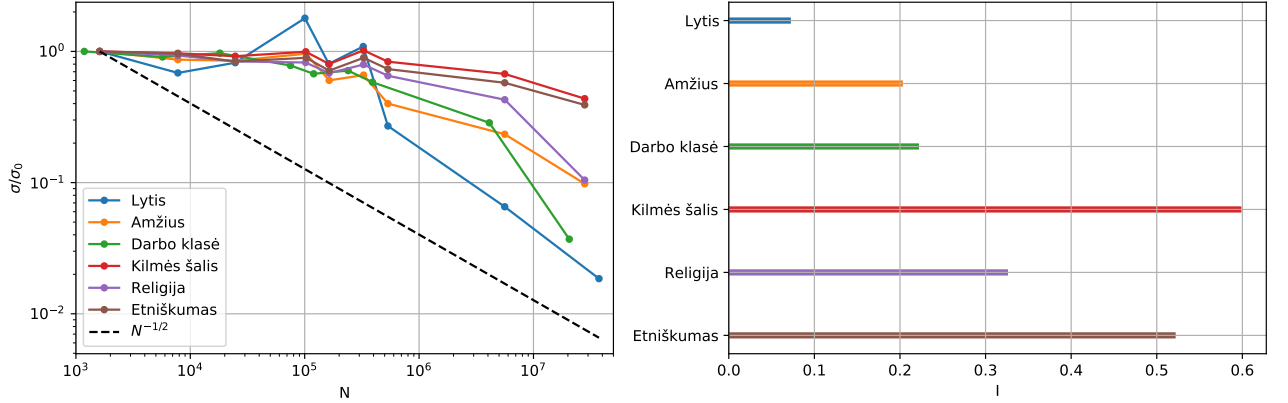


10 pav. Theil indekso verčių kitimas keičiant duomenų mastelį. T/T_0 žymi santykinę indekso vertę (lyginant su smulkiausio mastelio verte), N žymi mastelio vidutinį skirstiniui tenkantį gyventojų skaičių. Punktyrinėmis linijomis pažymėtas atsitiktinio modelio kreivės 95% patikimumo intervalas.

Iki šiol empiriniams duomenims buvo aptartas įvairovę nusakančių dydžių kitimas. Tačiau kiekybinei analizei yra reikalingas įvairovės indekso nustatymas. Tai galima padaryti pasinaudojant 8 paveiksle pavaizduotomis empirinių duomenų santykinio standartinio nuokrypio kitimo kreivėmis.

Standartinio nuokrypio verčių kitimas skirtingiems demografiniams rodikliams pateiktas 11 paveiksle. Kartu šiame paveiksle pavaizduotos ir demografiniams rodikliams gautos indekso vertės. Mažiausiomis indekso vertėmis pasižymi lyties ir amžiaus demografiniai rodikliai, o didžiausiomis - etniškumas bei kilmės šalis. Sudėtinga teigti ar erdvinė struktūra egzistuoja pagal lytį arba amžių, gyvenamosios vietos atskirtis pagal šiuos kriterijus nėra sutinkamas reiškinys. Kita vertus getų formavimasis ir segregacija yra žinomi reiškiniai, kurie sociofizikoje

nagrinėjami remiantis Šelingo modeliu.



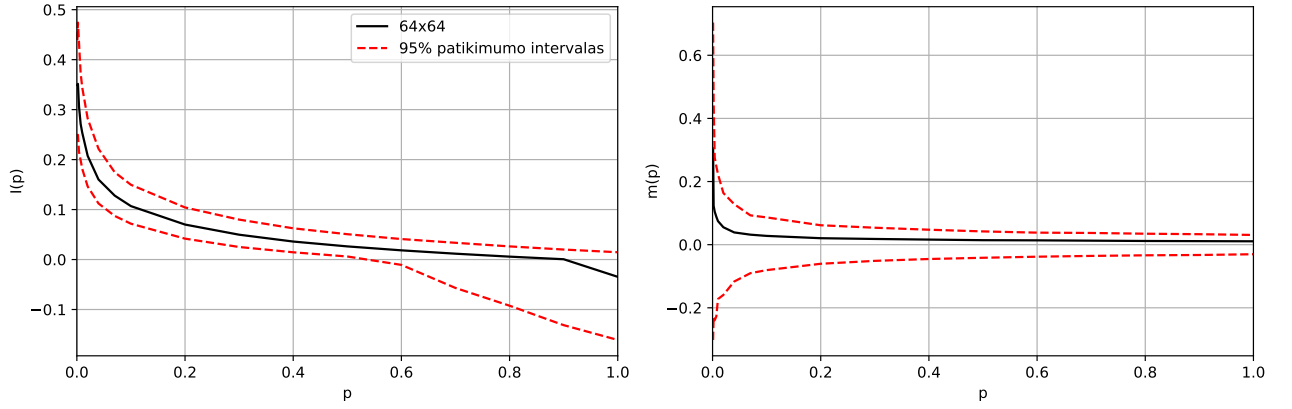
11 pav. Skirtingų demografinių rodiklių standartinio nuokrypio verčių kitimas keičiant empirinių duomenų mastelį (kairėje) ir jų indekso verčių palyginimas (dešinėje). I žymi įvairovės indekso vertės, σ/σ_0 žymi santykinės standartinio nuokrypio vertės, o N žymi vidutinį mastelio erdvianiam vnt. tenkanti gyventojų skaičių. Punktyrinė linija žymi atsitiktinio modelio rezultatams galiojanti dėsnį.

4.2 Įvairovės indekso verčių kitimas rinkėjo modeliui su triukšmu

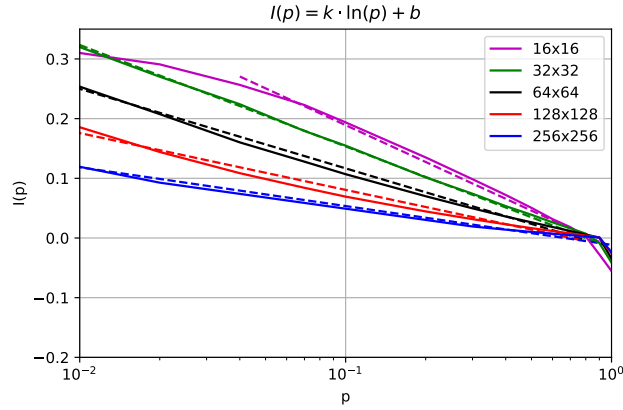
Nagrinėjant įvairovės indekso verčių kitimą teoriniuose modeliuose iš pradžių buvo išanalizuotas rinkėjo su triukšmu modelis. Triukšmas buvo įgyvendintas pagal 2.2.1 skyriuje aprašytą metodiką. Triukšmą nustatantis parametras įgyjo vertes intervale $(0; 1]$ ir atitinkamai nulemdavo tikimybę sukiniui keistis nepriklausomai nuo kaimyninių sukinių. Taip pat atliekant Monte Carlo simuliacijas gardelei buvo pritaikytos periodinės kraštinės sąlygos.

Siekiant išvengti sistemos užstrigimo lokaliuose minimumuose, kiekviena simuliacija su atitinkama triukšmo verte buvo pradedama konfigūracijoje, kurioje visi sukiniai yra vienodų verčių. Iš pradžių įgyvendinta 2 tūkst. ciklų, iš kurių kiekviename yra L^2 galinčių atsitiktinio sukinių vertę pakeisti įvykių. Čia L žymi kvadratinės gardelės kraštinėje esančių sukinių skaičių. Po sistemos įšildymo per 8 tūkst. ciklų yra įvertinamos įvairovės indekso ir vidutinės magnetizacijos vertės, atitinkamai apskaičiuotas 95% patikimumo intervalas ir dydžio medianos.

Įvairovės indekso ir magnetizacijos verčių priklausomybė nuo triukšmo dydžio pavaizduotos 12 paveiksle. Vertės gautos modeliuojant gardele, kuriai $L = 64$. Įvairovės indeksui matyti, jog jo vertė didžiausia, kai triukšmas mažiausias. Jeigu rinkėjo modelio konfigūracija pasiekia konsensuso būseną, tuomet indekso vertė tampa lygi vienetui. Ši būseną atitinka poliarizacijos atvejį tik su vienos vertės sukiniiais. Didėjant triukšmui stebimas staigus indekso vertės mažėjimas, jo vertė yra artima nuliui, o tai reiškia, jog sistema su triukšmu būdama pusiausvyroje yra atsitiktinėje konfigūracijoje. Triukšmo vertėms artimoms $p = 1$, įvairovės indeksas įgyja neigiamas vertes. Didelėms triukšmo vertėms daugiausiai vyksta atsitiktinis sukinių verčių ki-



12 pav. Įvairovės indekso I (kairėje) ir magnetizacijos m (dešinėje) verčių medianos priklausomybė nuo triukšmo. Dydis p žymi triukšmo tikimybę. Rezultatai gauti gardeli, kurios dydis yra 64x64. Raudona brūkšniuota linija žymi 95% verčių intervalą.



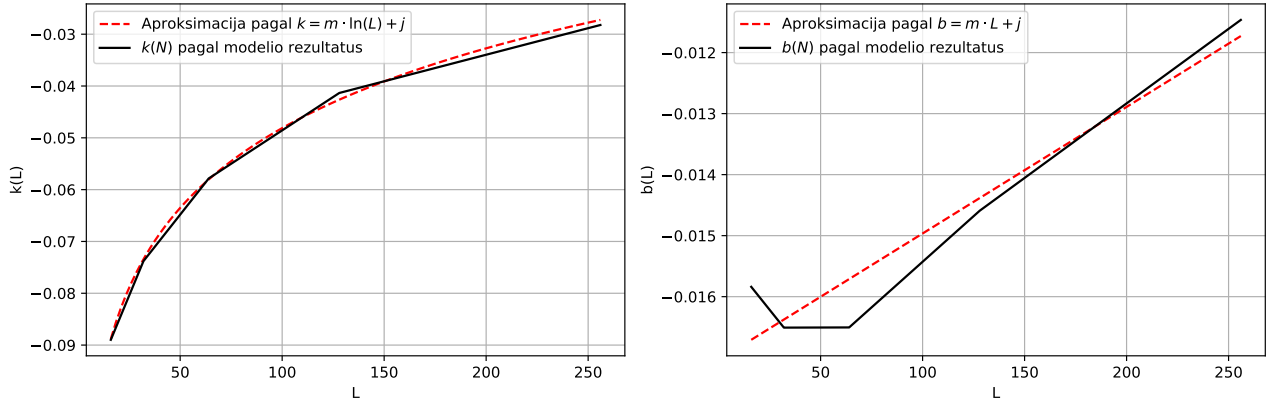
13 pav. Įvairovės indekso I verčių medianos priklausomybė nuo triukšmo p logaritinėje skalėje. Skirtingos tiesės žymi legendoje apibrėžtas gardeles. Logaritinė p verčių ašis parodo, jog įvairovės indekso priklausomybė nuo triukšmo yra logaritinė.

timas, o proporcinė rinkėjo modelio taisyklė nesukuria didesnių vienkrypčių sukinių domenų. Tai reiškia, jog lokaliai atsiranda tvarkingesnis išsidėstymas.

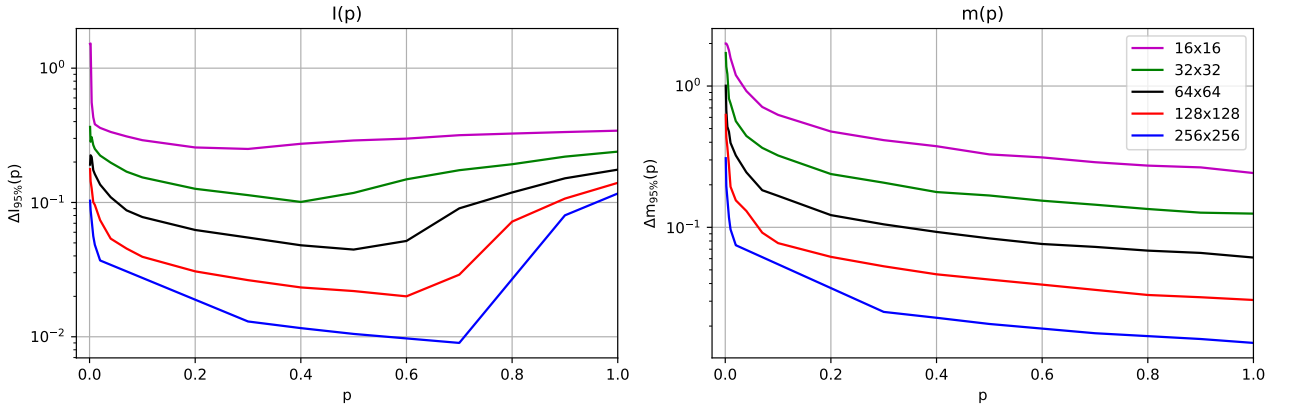
Magnetizacijos verčių kitimui reikia pastebėti, jog 12 paveiksle atvaizduota absoliučių magnetizacijos verčių mediana. Ji nėra pilnai simetrinė ašies $m = 0$ atžvilgiu, kadangi skaičiuojamos m modulio vertės. 95% patikimumo intervalas yra beveik simetriškas minėtos ašies atžvilgiu bei jo absoliutus plotis mažėja kai triukšmo vertė didėja.

Įvertinus įvairovės indekso vertes skirtingų dydžių gardelėms galima pastebėti, jog kitimas nuo triukšmo vyksta pagal tą patį dėsnį. Indekso vertės kinta pagal natūraliojo logaritmo dėsnį ir priklauso nuo gardelės dydžio. To aproksimacija yra pateikiama 13 paveiksle. Šios aproksimacijos koeficientų k ir b priklausomybės nuo gardelės dydžio pateikiamos 14 paveiksle.

Nuo gardelės dydžio priklauso ne tik nagrinėjamų tvarkos dydžių medianos, bet ir verčių patikimumo intervalo plotis. Pločio priklausomybės magnetizacijai ir įvairovės indeksui nuo



14 pav. Įvairovės indekso verčių aproksimacijos koeficientų k ir b priklausomybės nuo gardelės dydžio. Kairėje atvaizduota pagal 13 paveikslą legendoje pateikiamos lygties koeficientas k , o dešinėje koeficientas b . Atitinkamai legendose pateiktos funkcijų, kuriomis aproksimuoti koeficientai, išraiškos. Dydis L žymi kvadratinės gardelės kraštinėje esančių sukinių skaičių (viso sukinių - L^2).

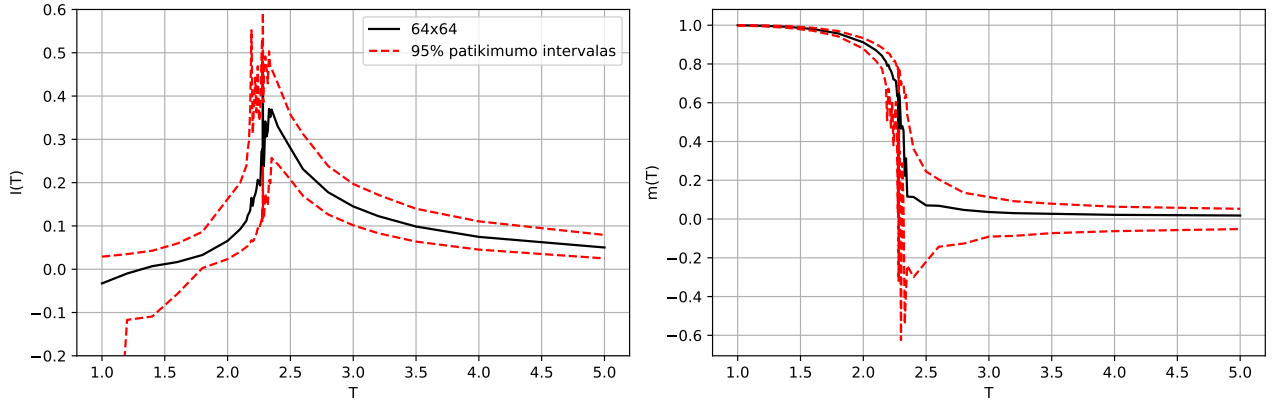


15 pav. Įvairovės indekso $\Delta I_{95\%}$ (kairėje) ir magnetizacijos $\Delta m_{95\%}$ (dešinėje) 95% patikimumo intervalo pločio priklausomybė nuo triukšmo. Logaritinė ordinačių ašis iliustruoja santykinį pločio skirtumą tarp verčių skirtingiems gardelių dydžiams.

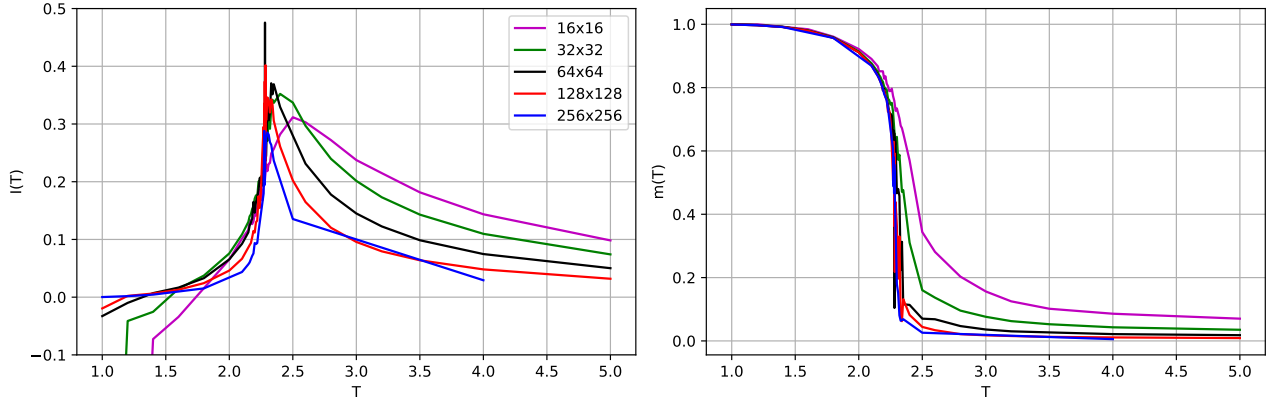
triukšmo yra pateikiamos 15 paveiksle. Dvigubinant abiejų gardelės kraštinių ilgi (4 kartus didinant sukinių skaičių), logaritinėje skalėje intervalo pokytis yra vienodas. Tai reiškia, jog santykinis pokytis yra eksponentinio pobūdžio. Skirtumas tarp patikimumo intervalo šiems dviem dydžiams yra tas, jog įvairovės indeksui dideliame triukšme stebimas intervalo išplitimas į neigiamą indekso verčių pusę.

4.3 Įvairovės indekso verčių kitimas Izingo modeliui pagal Metropolisio interpretaciją

Izingo modelio įgyvendinimas buvo atliktas pagal 2.2.2 skyriuje aptartą Metropolisio algoritmą. Simuliacijoje naudoti panašūs parametrai kaip ir rinkėjo su triukšmu modeliui - pusiausvyros būsenai pasiekti skirta 2 tūkst. ciklų, o verčių medianos ir patikimumo intervalo nustatymui



16 pav. Įvairovės indekso I (kairėje) medianos ir magnetizacijos modulio m (dešinėje) medianos verčių kitimo dėsniai Izingo modeliui pagal Metropolio interpretaciją. Vertės gautos gardelėi, kurios sukinių skaičius vienoje gardelės kraštinėje $L = 64$. Dydis T žymi santykinę temperatūrą, kuri yra normuota į Bolcmano ir feromagnetinės sąveikos konstantas.



17 pav. Įvairovės indekso I (kairėje) medianos ir magnetizacijos modulio m (dešinėje) medianos verčių kitimo dėsniai Izingo modeliui pagal Metropolio interpretaciją. Skirtingos kreivės žymi vertes gautas skirtingo dydžio gardelėms. Dydis T žymi santykinę temperatūrą, kuri yra normuota į Bolcmano ir feromagnetinės sąveikos konstantas.

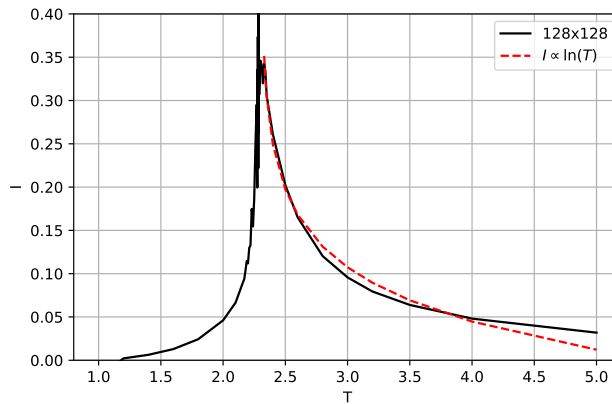
- 8 tūkst. ciklų. Gardelės pradinė konfigūracija būdavo parenkama šaltoje būsenoje, t.y. sukinių vertės būdavo vienodos.

Izingo modeliui įvairovės indekso ir magnetizacijos verčių dinamika yra skirtinga negu prieš tai nagrinėto rinkėjo modelio su triukšmu. Magnetizacija žemiau kritinio taško įgydavo nenulinę vertę, taško aplinkoje pasižymėjo gesimu pagal laipsninę funkciją, o paramagnetinėje būsenoje jos vertė apytikriai lygi nuliui. Šiuos dėsningumus galima matyti atitinkamai 16 paveiksle (dešinėje) su 95% večių intervalu bei 17 paveiksle (dešinėje) skirtingiems gardelių dydžiams. Atitinkamai tai buvo pastebėta prieš tai 3 paveiksle.

Įvairovės indeksas tuo tarpu įgyjo nenulines vertes kritinio taško aplinkoje ir pasižymėjo kitokia dinamika negu rinkėjo modelio atveju. Izingo modeliui literatūros apžvalgoje buvo pastebėta, kad šiame taške pasireiškia skalės invariantiškumas. Tai reiškia, jog keičiant mastelį,

arba stebėjimo skyrą, domenų dydžiai, kurie šiame taške būna įvairiausių išmatavimų, nesikeičia [1]. Todėl skaičiuojant įvairovės indeksą gaunamas ne atsitiktinis sukinių pasiskirstymas, o vertė yra artimesnė poliarizacijos atvejui. Šie dėsningumai pavaizduoti 16 ir 17 paveiksluose.

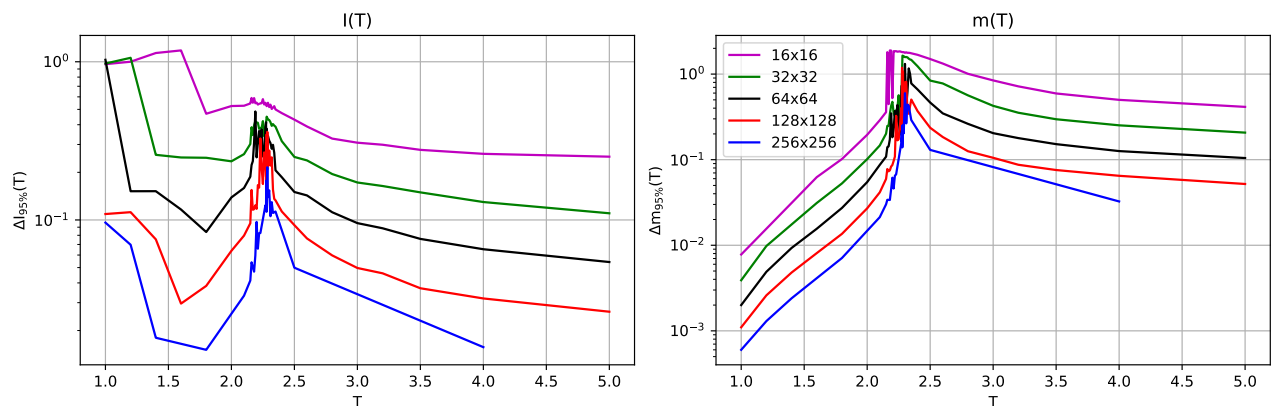
Indekso medianos verčių kitimas šio taško aplinkoje pasižymi dar keliais dėsningumais. Pirmiausia, smulkesniems gardelės dydžiams didžiausia indekso vertė gaunama didesnėms temperatūroms negu yra žinoma kritinio taško vertė. Todėl reikalingos didesnės gardelės norint tokiu būdu išmatuoti kritinio taško vertę. Taip pat reikia pastebėti, jog įvairovės indekso verčių kitimas nėra simetriškas kritinio taško atžvilgiu. Didesnėms temperatūroms galima matyti lėtesnį indekso verčių mažėjimą negu kylimą mažesnėms temperatūroms. Be to, šis mažėjimas vyksta pagal natūraliojo logaritmo dėsnio, kuris buvo stebėtas rinkėjo modelio su triukšmu atveju. Aproximacija pateikiama 18 paveiksle. Skirtingų dydžių gardelėms šis kitimas matomas 17 paveiksle (kairėje).



18 pav. Įvairovės indekso I verčių kitimo dėsnis nuo T Izingo modeliui pagal Metropolio interpretaciją už kritinio temperatūros taško. Kritinio temperatūros taško vertė $T_C = 2.269 T_0$. Punktyrinėmis linijomis pažymėtos aproksimacijos dėsnis yra įvardintas paveikslo legendoje.

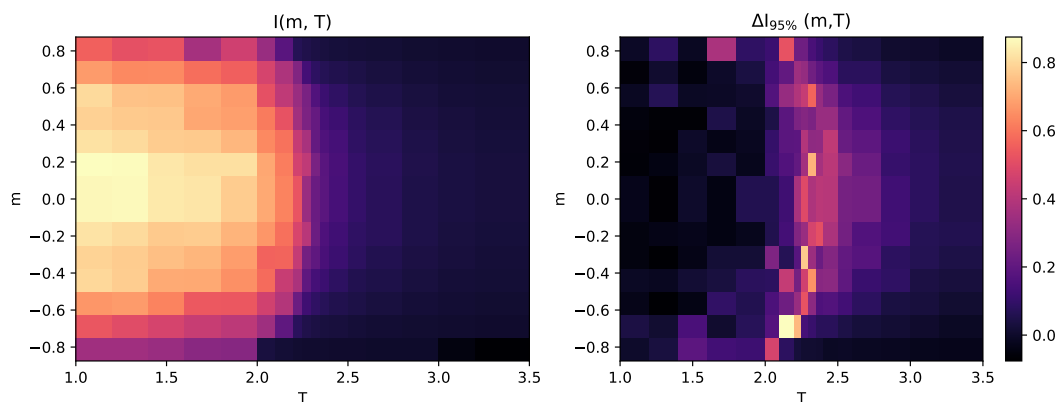
Vis dėlto įvairovės indekso verčių kitimas yra priešingas negu intuityviai būtų tikimasi žemų temperatūrų srityje. Žemesnėse negu kritinio taško temperatūrose Izingo modelyje susiformuoja vienas didelis domenai ir keletas smulkesnių. Tai reiškia, jog tam tikra tvarka susidaro. Vis dėlto įvairovės indekso vertėse tai nėra užfiksuojama. Tai gali būti susiję su tuo, jog naudojama ganėtinai paprasta erdvinių vienetų jungimo procedūra, kuri ignoruoja besiformuojančių domenų padėtis ir formas.

Magnetizacijos ir indekso verčių 95% intervalo dėsningumai yra panašesni į gautus rinkėjo modelio atveju. Jie yra pavaizduoti 19 paveiksle. Logaritminėje verčių ašyje matyti, kad tarpai tarp skirtingų gardelės dydžių kreivių yra vienodi, o tai parodo, jog santykinis verčių mažėjimas yra eksponentinio pobūdžio. Vienas didesnių skirtumų lyginant Izingo modelio atveju 19 paveikslą su rinkėjo modelio 15 paveikslu yra tas, jog kritinio taško aplinkoje stebimas intervalo pločio padidėjimas. Indekso vertėms intervalo plotis yra labiau simetriškas negu magnetizacijai kritinio taško atžvilgiu.



19 pav. Įvairovės indekso $\Delta I_{95\%}$ (kairėje) ir magnetizacijos $\Delta m_{95\%}$ (dešinėje) 95% patikimumo intervalo pločio priklausomybė nuo triukšmo Izingo modeliui pagal Metroplio interpretaciją. Logaritminė ordinatės ašis iliustruoja santykinį pločio skirtumą tarp verčių skirtingiems gardelių dydžiams.

4.4 Įvairovės indekso verčių kitimas Izingo modeliui pagal Kawasaki interpretaciją



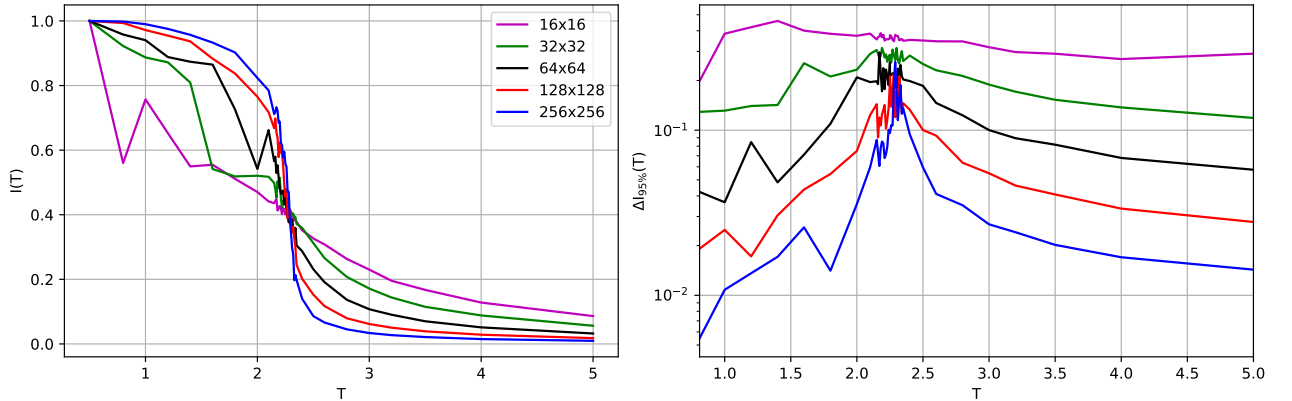
20 pav. Įvairovės indekso I verčių medianos priklausomybė nuo temperatūros T ir magnetizacijos m (kairėje) Izingo modeliui pagal Kawasaki interpretaciją. Dešinėje pavaizduota 95% verčių patikimumo srities dydžio $\Delta I_{95\%}$ priklausomybė nuo tų pačių parametrų. Rezultatai gauti gardelei, kurios dydis yra 128x128.

Nagrinėjant Izingo modelį pagal Kawasaki interpretaciją buvo pasirinktos tos pačios modeliavimo sąlygos kaip ir prieš tai minėtiems modeliams. Pagrindinė šio modelio ypatybė - nesikeičiantis vienkrypčių sukinių skaičius - lemia tai, jog kiekvienai simuliacijai parenkami du nekintantys parametrai - temperatūra ir magnetizacija. Taip pat atliekant modeliavimus sukinių sukeitimas buvo atliekamas globaliai. Tai reiškia, jog sukiniai vietomis keitėsi ne tik su kaimynais, tačiau su bet kuriuo priešingos vertės sukiniu gardelėje. Tai lemia kitokią dinamiką perėjime į pusiausvyros būseną, tačiau įvairovės indekso verčių skaičiavimui pusiausvyros būsenoje poveikio tai neturi.

Derinant temperatūros ir magnetizacijos verčių parinkimą galima įvairovės indekso priklausomybę pavaizduoti fazinėje erdvėje. Ši priklausomybė yra pateikiama 20 paveiksle (kairėje). Įvairovės indekso vertės yra didelės žemų temperatūrų srityje ir artima nuliui aukštesnėse negu kritinio temperatūros taško srityje. Skirtingai nuo Izingo modelio pagal Metropolį kritinio taško aplinkoje įvairovės indeksas tik mažėja. Šis mažėjimas yra panašesnis į laipsninį magnetizacijos kitimą modeliui pagal Metropolio interpretaciją. Tai vizualiai matyti 21 paveiksle (kairėje) Šio kitimo statusas taip pat priklauso nuo modeliujamos gardelės dydžio, t.y. smulkesnėms gardelėms jis yra mažiau status.

Žvelgiant į paveikslą ir laikant temperatūras pastoviomis matoma kitokia dinamika. Žemų temperatūrų srityje didžiausios indekso vertės yra tuomet kai magnetizacija artima nuliui. Artėjant prie ribinių magnetizacijos verčių indekso vertės palyginimai sumažėja. Tačiau temperatūros sričiai virš kritinio taško skirtingos magnetizacijos vertės turi ženkliai mažesnę poveikį įvairovės indekso vertėms.

Fazinės erdvės paveikslo dešinėje galima matyti absoliutaus patikimumo juostos pločio $\Delta I_{95\%}$ priklausomybę nuo parametrų m ir T . Šio vertės yra didžiausios kritinio temperatūros taško srityje. Patikimumo juostos pločio vertės padidėjimas kritinio taško srityje taip pat buvo matomas Izingo modeliui 19 paveiksle (kairėje). Kaip ir prieš tai nagrinėtų sukininių modelių atveju, patikimumo juostos intervalo plotis mažėja eksponentiškai didinant sukinių skaičių gardelėje. Skirtingiems gardelės dydžiams tai pavaizduota 21 paveiksle (dešinėje).



21 pav. Įvairovės indekso I verčių medianos ir 95% verčių patikimumo srities dydžio $\Delta I_{95\%}$ priklausomybės nuo temperatūros T . Skirtingos spalvos kreivės žymi rezultatus gautus įvairaus dydžio gardelėms.

Iš fazinės diagramos paveiklo galima pastebėti, jog kritinės temperatūros sritis nežymiai pasislenka artimoms ± 1 magnetizacijos vertėms. Pagal [15] tai nutinka dėl to, jog fiksuotas sukinių populiacijų dydis save faziškai atskiria (angl. *phase separation*) ir suformuoja sritis, kuriose teigiamų sukinių tankis ρ lokaliai turi dvi pusiausvyros vertes:

$$\rho_{\pm} = \frac{1}{2}(1 \pm |m|), \quad (29)$$

kur $|m|$ žymi magnetizacijos absoliučias vertes. Riba tarp šios srities ir atsitiktinio pasiskirstymo sričių nuo magnetizacijos įgyja ne tiesės formą, o užlinksta (sumažėja temperatūrų ašyje) kraštinės magnetizacijos atvejams. Taigi šių dviejų skirtingų tankių koegzistavimo srityje įvairovės indeksas įgyja didžiausias vertes.

5 Išvados

1. Visiems empirinių duomenų demografiniams rodikliams nustatytas lėtesnis įvairovę nusakančių dydžių santykinis mažėjimas lyginant su atsitiktinio modelio atveju gautais porcingumais. Kiekybiškai reikšmingiausią nukrypimą nuo atsitiktinio modelio turi kilmės šalies ir etniškumo rodikliai, jų indekso vertės artimesnės poliarizacijos atvejui negu atsitiktiniam pasiskirstymui.
2. Izingo modelio pagal Metropolio interpretaciją paramagnetinėje srityje ir rinkėjo su triukšmu modelyje pasiūlyto įvairovės indekso vertės mažėja kai dažnėja savaiminis, nepriklausantis nuo aplinkos, sukinio apsisvertimas. Indekso vertės nuo temperatūros ir triukšmo slopsta pagal natūraliojo logaritmo dėsnį.
3. Sukininių sistemų gardelių baigtinis dydis turi įtakos įvairovės indekso verčių išsisklaidymui. Indekso 95% patikimumo intervalo plotis mažėja eksponentiškai didinant modeliuojamos gardelės sukinų skaičių.
4. Izingo modeliui Metropolio interpretacijos atveju didžiausios įvairovės indekso vertės (artimos 0.4) buvo stebimos kritinės temperatūros aplinkoje. Kawasaki interpretacijos atveju didžiausios įvairovės indekso vertės (artimos 1) buvo stebimos žemų temperatūrų aplinkoje, domenų koegzistavimo fazėje.
5. Empiriniams duomenims stebėtos įvairovės indekso vertės atitinka indekso vertes stebėtas Izingo modelyje kritinės temperatūros aplinkoje. Geriausias atitikimas stebėtas Izingo modelio Kawasaki interpretacijai. Šis rezultatas leidžia teigti, kad Kawasaki interpretacija yra tinkamiausia demografiniams reiškiniams modeliuoti.

ERDVINIŲ SKIRSTINIŲ POKYČIAI KINTANT EMPIRINIŲ DUOMENŲ MASTELIUI

Santrauka

Tarpdisciplininių tyrimų sritis vadinama sociofizika tyrinėja sudėtingus vyksmus ir reiškinius socialinėse sistemose tam pasitelkdama statistinės fizikos metodus. Nuomonių dinamika yra vienas plačiausiai tirtų reiškinių šioje srityje. Įvairioms rinkėjų nuomonių apsikeitimo dinamikoms buvo pasiūlyti skirtingi modeliai, iš kurių žinomiausias ir paprasčiausias yra rinkėjo modelis. Jis panašus į statistinėje fizikoje plačiai ištirtą ir sąlyginai paprastą Izingo modelį.

Vis dėlto tyrėjai kelia klausimą - kiek rinkėjo modelis yra tinkamas tirti realių žmonių nuomonių dinamika? Įprasta šiuo modeliu simuliuojama nuomonių laikinė dinamika neįvertina to, kad žmonės ir bendruomenės keičia savo vietą erdvėje, kasdien keliauja į darbą ar poilsio vietas. Atsižvelgdami į galimybę gauti ir išreikšti socialinį poveikį ne tik laikinėje dinamikoje, bet ir erdvinėje, sociofizikai pasiūlo išplėsti modelius įtraukiant nuo erdvės priklausančią dinamiką. Norint suprasti erdvės įtaką nuomonių dinamiką pradžioje reikalinga suprasti kaip empiriniai ar teoriniai duomenys gali parodyti nuomonių erdvines struktūras.

Taigi šio darbo tikslas - paaiškinti erdvinių skirstinių pokyčius kintant duomenų masteliui. Darbe yra apibendrinami pastebėjimai analizuojant Jungtinės Karalystės 2011 m. gyventojų surašymo duomenis. Pagal šiuos pastebėjimus analizuojami teoriniai sukininiai modeliai ir kaip erdviniai skirstiniai kinta keičiant mastelį juose. Erdvės struktūra įvertinama pasiūlant du įvairovės matavimo metodus - atsitiktinį erdvinių vienetų rikiavimo modelį ir skaitinį įvairovės indeksą. Pagrindiniai darbo rezultatai ir išvados yra:

1. Visiems empiriniais duomenimis apibūdinamiems demografiniams rodikliams nustatytas lėtesnis įvairovės nusakančių dydžių santykinis mažėjimas lyginant su atsitiktinio modelio atveju gautais proporcingumais. Kiekybiškai reikšmingiausią nukrypimą nuo atsitiktinio modelio turi kilmės šalies ir etniškumo rodikliai.
2. Izingo modelyje pagal Metropolio ir Kawasaki interpretacijas yra stebimos sritys, kuriose įvairovės indekso vertės artimos gardelės poliarizacijos konfigūracijai. Metropolio interpretacijos atveju didžiausios įvairovės indekso vertės (artimos 0.4) buvo stebimos kritinės temperatūros aplinkoje. Kawasaki interpretacijos atveju didžiausios įvairovės indekso vertės (artimos 1) buvo stebimos žemų temperatūrų aplinkoje, domenų koegzistavimo fazėje.
3. Empiriniais duomenimis stebėtos įvairovės indekso vertės atitinka įvairovės indekso vertes stebėtas Izingo modelyje kritinės temperatūros aplinkoje. Geriausias atitikimas stebėtas Izingo modelio Kawasaki interpretacijai. Šis rezultatas leidžia teigti, kad Kawasaki interpretacija yra tinkamiausia demografiniams reiškiniams modeliuoti.

SCALING OF EMPIRICAL COMPARTMENTAL DISTRIBUTIONS

Summary

Sociophysics is an interdisciplinary research discipline which studies social phenomena using tools from statistical physics. Opinion dynamics is the focal topic in sociophysics. Various opinion dynamics models have been proposed of which voter model with its variations is the most studied one. It is similar to a well-known Ising model in statistical physics.

One of the recent research questions in sociophysics is how appropriate voter model is for studying real social dynamics of opinion change? Usual temporal evolution of opinion exchange does not account for simple spatial effects - commuting, workplace or leisure events. Researchers have suggested to expand models for opinion dynamics by including spatially dependent dynamics. In order to understand the implications of spatiality for opinion dynamics it is necessary to understand how empirical or theoretical data can depict structures of space.

The goal of this bachelor thesis is to explain changes that happen in the compartmental distributions as the empirical data is examined at different scales. Particularly 2011 census data of UK was analyzed in this regard and results are presented. Further analysis encompasses theoretical spin models, namely voter model and Metropolis, Kawasaki interpretations of Ising model. Scaling effects are analyzed by developing two spatial diversity calculating measures - random spatial model and index of diversity.

Main results of the research are the following:

1. All demographic indicators from empirical data show slower change of diversity measures than suggested by random spatial model. In terms of diversity index values two indicators are the most diverse - country of origin and ethnicity.
2. In both interpretations of the Ising model we have observed phases with increased values of the diversity index. For Metropolis interpretation this phase was observed in the vicinity of critical temperature (with diversity index values close to 0.4). For Kawasaki interpretation this phase was observed for the lower temperatures in the domain coexistence phase (with diversity index values close to 1).
3. Values of the diversity index observed for the empirical data are similar to those observed near the critical temperature of the Ising model. Kawasaki interpretation seems to generate broadest range of possible values. This result indicates that Kawasaki interpretation is the most applicable in understanding spatial social phenomena.

Literatūros sąrašas

- [1] B. K. C. Parongama Sen, *Sociophysics: An Introduction* (Oxford university Press, 2014)
- [2] C. Castellano, S. Fortunato, V. Loreto, Statistical physics of social dynamics, *Rev. Mod. Phys.* **81**, 591–646 (2009), <http://dx.doi.org/10.1103/RevModPhys.81.591>
- [3] J. Fernández-Gracia, K. Suchecki, J. J. Ramasco, M. Miguel, V. Eguíluz, Is the voter model a model for voters?, *Physical Review Letters* **112**, 158701 (2014), <http://dx.doi.org/10.1103/PhysRevLett.112.158701>
- [4] A. Kononovicius, Compartmental voter model, *Journal of Statistical Mechanics: Theory and Experiment* **2019**, 103402 (2019), <http://dx.doi.org/10.1088/1742-5468/ab409b>
- [5] J. Kruopis, *Matematinė statistika* ("Mokslo" leidykla, 1993)
- [6] S. So, Why is the sample variance a biased estimator?, https://maxwell.ict.griffith.edu.au/ssso/biased_variance.pdf, (Žiūrėta 2022 m. gegužės 16 d.) (2008)
- [7] T. Van Ourti, P. Clarke, A simple correction to remove the bias of the gini coefficient due to grouping, *The Review of Economics and Statistics* **93**, 982–994 (2011), <http://dx.doi.org/10.2307/23016090>
- [8] S. K. Milorad, A. B. David, Variance estimation for measures of income inequality and polarization - the estimating equations approach, *Journal of Official Statistics* **13**(1), 41, copyright - Copyright Statistics Sweden (SCB) Mar 1997; Last updated - 2013-01-07 (1997), <https://www.proquest.com/scholarly-journals/variance-estimation-measures-income-inequality/docview/1266835912/se-2>
- [9] R. Dorfman, A formula for the gini coefficient, *The Review of Economics and Statistics* **61**(1), 146–149 (1979), <http://www.jstor.org/stable/1924845>
- [10] F. Maio, Income inequality measures, *Journal of epidemiology and community health* **61**, 849–52 (2007), <http://dx.doi.org/10.1136/jech.2006.052969>
- [11] F. Cowell, *Measurement of inequality* (2000), 87–166, [http://dx.doi.org/10.1016/S1574-0056\(00\)80005-6](http://dx.doi.org/10.1016/S1574-0056(00)80005-6)
- [12] M. Langel, Y. Tillé, Variance estimation of the gini index: revisiting a result several times published, *Journal of the Royal Statistical Society. Series A (Statistics in Society)* **176**(2), 521–540 (2013), <http://www.jstor.org/stable/23355203>
- [13] P. Conceição, P. Ferreira, A young person's guide to the theil index: Suggestive intuitive interpretations and exploring analytical applications, *SSRN Electronic Journal* (2000), <http://dx.doi.org/10.2139/ssrn.228703>

- [14] D. Stauffer, A biased review of sociophysics, *Journal of Statistical Physics* **151** (2012), <http://dx.doi.org/10.1007/s10955-012-0604-9>
- [15] M. E. J. Newman, G. T. Barkema, *Monte Carlo Methods in Statistical Physics* (Clarendon Press, Oxford, 1999)
- [16] P. Krapivsky, S. Redner, E. Ben-Naim, *A kinetic view of statistical physics*, Cambridge, UK: Cambridge University Press, 2010 349 (2010), <http://dx.doi.org/10.1017/CB09780511780516>
- [17] S. Mori, M. Hisakado, K. Nakayama, Voter model on networks and the multivariate beta distribution, *Physical Review E* **99**(5) (2019), <http://dx.doi.org/10.1103/physreve.99.052307>
- [18] H. Jang, M. Grimson, T. Woolf, Stochastic dynamics and the dynamic phase transition in thin ferromagnetic films, *Physical review. E, Statistical, nonlinear, and soft matter physics* **70**, 047101 (2004), <http://dx.doi.org/10.1103/PhysRevE.70.047101>
- [19] S. Galam, Sociophysics: A review of galam models, *International Journal of Modern Physics C* **19** (2008), <http://dx.doi.org/10.1142/S0129183108012297>
- [20] T. Fenner, E. Kaufmann, M. Levene, G. Lozou, A multiplicative process for generating a beta-like survival function with application to the uk 2016 eu referendum results, *International Journal of Modern Physics C* **28** (2017), <http://dx.doi.org/10.1142/S0129183117501327>