



**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

BAKALÁŘSKÁ PRÁCE

František Miroslav Škola

**Symetrie Stock Network založené na
automorfismech**

Informatický ústav Univerzity Karlovy

Vedoucí bakalářské práce: doc. Ing. et Ing. David Hartman,
Ph.D. et Ph.D.

Studijní program: Informatika

Praha 2024

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V dne

Podpis autora

Děkuji vedoucímu práce doc. Ing. et Ing. Davidu Hartmanovi, Ph.D. et Ph.D. za ochotu, podporu a inspiraci během celého vypracování práce. Také bych chtěl poděkovat své rodině, která mě ve studiích vždy podporovala.

Název práce: Symetrie Stock Network založené na automorfismech

Autor: František Miroslav Škola

Ústav: Informatický ústav Univerzity Karlovy

Vedoucí bakalářské práce: doc. Ing. et Ing. David Hartman, Ph.D. et Ph.D.,
Informatický ústav Univerzity Karlovy

Abstrakt: Komplexní sítě se využívají k modelování systémů napříč různými obory, mimo jiné v několika oblastech ekonomie, konkrétně pro oblast obchodování na akciových trzích. Jednou z konkrétních vlastností sítí, která bývá častým předmětem zájmu při jejich zkoumání je jejich symetrie. Jedna z možností vyjádření symetrie jsou různé charakteristiky automorfnní grupy. V této práci zkoumáme vlastnosti takové symetrie pro sítě tvořené nad daty akciových trhů, tzv. Stock Network, a jejich propojení s jevy reálného světa.

Klíčová slova: komplexní sítě, Stock Network, symetrie, automorfismus

Title: Symmetry of Stock Network based on automorphisms

Author: František Miroslav Škola

Institute: Computer Science Institute of Charles University

Supervisor: doc. Ing. et Ing. David Hartman, Ph.D. et Ph.D., Computer Science
Institute of Charles University

Abstract: Complex networks are used to model accross differenet disciplines, among others several areas of economics, specifically in the area of stock market. One of the many properties of networks that is often a subject of interest is their symmetry. Possible ways to express symmetry are different properties of the automorphism group. In the thesis, we explore properties of such symmetry of networks based on data from stock markets — Stock Networks — and its connection to events of the real world.

Keywords: complex networks, Stock Network, symmetry, automorphisms

Obsah

Úvod	6
1 Základní definice	8
1.1 Několik pojmů z teorie grafů	8
1.2 Automorfismus a symetrie	9
1.3 Sítě a jejich vlastnosti	9
1.4 Korelace a vzájemná informace	10
2 Stock network	12
2.1 Data	12
2.2 Parametry dat, jejich získání a zpracování	12
2.3 Parametry sítě a sestavení grafu sítě	13
2.4 Získávání komponenty souvislosti	15
2.5 Měřené veličiny	18
2.6 Obarvení grafu podle GICS	18
2.7 Použitý software	19
3 Analýza sítí	20
3.1 Metoda 1	20
3.2 Metoda 2	23
3.3 Metoda 3	25
3.4 Shrnutí	26
4 Možnosti nahrazení symetrie	30
4.1 Vlastnosti nesouvisející se symetrií	30
4.2 Weisfeiler-Lehmanův test	37
Závěr	45
4.3 Prostor pro další práci	45
Literatura	47
Seznam obrázků	49
Seznam tabulek	52

Úvod

Využití komplexních sítí v modelování reálných systémů není žádnou novinkou [1]. Zkoumání těchto modelů se ukazuje jako velice efektivní v odhalování strukturních vlastností reálných systémů, které modelují [2]. I proto se s nimi můžeme setkat napříč vědními obory — od medicíny [3] a biologie [4] po modelování internetu [5]. Konkrétně se sítě ve výzkumu využívají tak, že se analyzují jejich vlastnosti, jako je třeba klastrovací koeficient.

Jednou ze zajímavých vlastností sítí je symetrie. Ta se zkoumá hlavně v sítích používaných ve fyzice, kde se používá k identifikování neměnnosti pod nějakou transformací [6]. Lze ji také použít k redukci sítě a díky tomu snížení výpočetní náročnosti výzkumu na ní [2]. Jako velikost automorfnní grupy na grafu sítě zavedl symetrii MacArthur [7]. Ten také ukázal, že mnoho sítí z reálného světa nějaké množství symetrií vykazuje [7], což bylo následně ověřeno i dalšími autory [2].

Symetrie nezůstala bez povšimnutí ani v ekonomii. V roce 2008 publikovali Wang, Yan a Xiao článek, ve kterém pomocí symetrií zkoumali strukturu mezinárodního trhu [8]. Představili zde síť obchodní bilance států a zjistili, že uzly reprezentující státy závislé na stejném obchodním partnerovi jsou symetrické, přičemž symetrií ve své práci mysleli právě výskyt automorfismů.

Devět let předtím, v roce 1999, vydal Mantegna článek [9], ve kterém představil Stock Network — síť, ve které uzly jsou firmy a vazby mezi nimi jsou ohodnoceny korelačním koeficientem časových řad cen jejich akcií. Ve svém textu se Mantegna zabýval maximální kostrou grafu reprezentujícím Stock Network a došel k tomu, že vrcholy reprezentující firmy ze stejného sektoru (zabývající se podobnou činností) spolu v této kostře sousedí.

V naší práci se zabýváme kombinací těchto dvou konceptů — symetrií Stock Network. Zajímá nás, jak moc je Stock Network symetrická a jestli lze míru její symetričnosti navázat na ekonomický cyklus. Víme, že akcie malého množství velkých firem mají velký dopad na celé trhy [10]. Ptáme se, jestli se tento dopad mění v časech kolapsů trhu — zaměřujeme se na světovou finanční krizi v letech 2007 a 2008, recesi způsobenou pandemií nemoci Covid-19 v roce 2020 a ochlazení ekonomik po vypuknutí války na Ukrajině v roce 2022.

Intuice nás vede k hypotéze, že dopad velkých firem na trh bude během časů nejistoty sílit, protože se menší firmy budou — nevědouce si rady — inspirovat u firem velkých. To se pak odrazí na cenách akcií, jejichž vývoj se na sebe začne vázat, což se zase odrazí v síti složené z akcií firem — větším výskytem symetrií.

Mantenga ve svém článku zkoumal maximální kostry grafů sítí. Přejít od grafu k jeho kostře, ale — obzvláště v případě Stock Network, jejíž graf je ve výchozím stavu úplný — způsobuje velký pokles počtu hran. Ke zkoumání symetrií se tedy nehodí. Mezi další používané přístupy získání zkoumané sítě patří mimo jiné maximální rovinný filtrovaný graf [11] a strategie vítěz bere vše [10].

Během naší práce jsme sestavili Stock Network z akcií pěti burzovních indexů a to jak pomocí korelačního koeficientu, tak pomocí odhadu vzájemné informace. Dále jsme navrhli 3 metody, jak z úplného grafu sítě získat souvislý graf s hustotou vhodnou pro zkoumání symetrií a zkusili jsme symetrii nahradit jinými, výpočetně méně náročnými charakteristikami, včetně odhadu symetrie pomocí Weisfeiler-Lehmanova testu.

Struktura naší práce je následující. V kapitole 1 zavedeme definice, které během práce budeme používat. V kapitole 2 zmíníme, proč nám běžně používané metody získání sítě nevyhovují a navrhneme 3 jiné metody, které v naší analýze použijeme. Kromě toho zde detailně popíšeme získání dat a postavení sítě a jejího grafu. V kapitole 3 představíme a popíšeme nejlepší výsledky. V kapitole 4 se podíváme na možnost nahrazení symetrie vlastnostmi jiného charakteru, konkrétně průměrnou nejkratší cestou a klastrovacím koeficientem. Dále se podíváme na odhad symetrie pomocí Weisfeiler-Lehmanova testu.

1 Základní definice

1.1 Několik pojmů z teorie grafů

Tato podkapitola čerpá ze základního kurzu diskrétní matematiky na MFF (NDMI002) a z Matouška a Nešetřila [12]

Definice 1 (neorientovaný graf). *Neorientovaný graf G je uspořádaná dvojice $G = (V, E)$, kde V je množina vrcholů a E množina hran: $E \subseteq \{\{a, b\} \mid a, b \in V\}$.*

V tomto textu budeme pracovat pouze s jednoduchými grafy bez smyček, tj. nedovolujeme hrany $\{a, b\}$, kde $a = b$ a mezi každými 2 vrcholy a, b vyžadujeme maximálně 1 hranu.

Definice 2 (cesta v grafu). *Cesta v grafu G je posloupnost vrcholů spojených hranou, ve které se každý vrchol a každá hrana vyskytuje maximálně jednou.*

Definice 3 (graf bez hrany). *Nechť $G = (V, E)$ je graf a e je jeho hrana — tedy $e \in E$, pro přehlednost budeme psát $e \in G$. Nechť G' je graf G bez hrany e . Píšeme:*

$$G' = G \setminus e$$

Definice 4 (souvislý graf). *Řekneme, že graf $G = (V, E)$, je souvislý, pokud mezi každými 2 různými vrcholy $a, b \in V$ existuje nějaká cesta.*

Definice 5 (komponenta souvislosti). *Komponenta souvislost H grafu G je jeho souvislý podgraf, tj. $H = (V', E')$, kde $V' \subseteq V$, $E' \subseteq E \cap V'$ a H je souvislý.*

V naší práci budeme pracovat hlavně s pojmem největší komponenta souvislosti. To je taková komponenta souvislosti, která má nejvíce vrcholů ze všech komponent souvislosti daného grafu. Největší komponentu souvislosti grafu G budeme značit $Giant(G)$.

Definice 6 (nejkratší cesta). *Nechť u, v jsou vrcholy v souvislém grafu G . Řekneme, že nejkratší cesta mezi u a v je taková, která má ze všech cest mezi u a v nejméně hran.*

Definice 7 (sousedství vrcholu). *Nechť v je vrchol v grafu G . Jako sousedství vrcholu v označujeme množinu všech vrcholů, které jsou s v spojeny nějakou hranou. Tuto množinu budeme značit $N(v)$.*

Definice 8 (stupeň vrcholu). *Nechť G je neorientovaný graf a v je jeho vrchol. Stupeň vrcholu v je počet hran z něj vedoucích. Značíme $deg(v)$.*

Protože pracujeme pouze s jednoduchými grafy — hrany vedou pouze mezi různými vrcholy a mezi každými 2 vrcholy existuje maximálně jedna hrana — můžeme říct, že stupeň vrcholu je roven velikosti jeho sousedství:

$$deg(v) = |N(v)|$$

1.2 Automorfismus a symetrie

Tato podkapitola čerpá z článku o symetriích v reálných sítích [13].

Definice 9 (izomorfismus a automorfismus grafu). *Mějme grafy $G = (V, E)$ a $H = (V', E')$. Zobrazení $f : V \rightarrow V'$ označíme jako izomorfismus, pokud je bijektivní a zachává strukturu, tedy platí: $\{a, b\} \in E \iff \{f(a), f(b)\} \in E'$. Pokud navíc $G = H$, pak řekneme, že f je automorfismus.*

Definice 10 (automorfní grupa grafu). *Množina všech automorfismů grafu G tvoří jeho automorfní grupu $\text{Aut}(G)$, tedy splňuje axiomy grupy:*

1. *Identita:* $1 \in \text{Aut}(G)$: *identické zobrazení je izomorfismus grafu G*
2. *Inverz:* $(\forall f \in \text{Aut}(G)) (\exists! f^{-1} \in \text{Aut}(G)) ff^{-1} = f^{-1}f = 1$
3. *Uzavřenost:* $(\forall f, g \in \text{Aut}(G)) fg \in \text{Aut}(G)$
4. *Asociativita:* $(\forall f, g, h \in \text{Aut}(G)) (fg)h = f(gh)$

Je-li $1 \in \text{Aut}(G)$ jediným prvkem automorfní grupy grafu G , pak řekneme, že je G asymetrický.

Definice 11 (orbita automorfní grupy). *Nechť $\text{Aut}(G)$ je automorfní grupa grafu $G = (V, E)$ a nechť $a \in V$ je vrchol G . Množinu všech vrcholů, na které se a zobrazuje nějakým automorfismem f : $[a] = \{f(a) \mid f \in \text{Aut}(G)\}$ nazýváme orbitou vrcholu a .*

Není těžké si uvědomit, že rozdělení vrcholů do orbit tvoří rozklad množiny V . Stačí se na automorfismus dívat jako na relaci. Tato relace je ve skutečnosti ekvivalencí a orbity tvoří její třídy [14].

1.3 Síť a jejich vlastnosti

Tato podkapitola čerpá ze základního kurzu o sítích na MFF (NDMI110).

Definice 12 (komplexní síť). *Jako komplexní síť označujeme netriviální strukturu sestávající z uzlů a vazeb mezi nimi.*

Co přesně se myslí tím, že je struktura netriviální? V případě sítí většinou mluvíme o následujících vlastnostech:

- Scalefree — to znamená, že zatímco většina uzlů je vazbami spojena jen s malým množstvím jiných, existuje několik uzlů, které mají vazeb na jiné uzly mnoho. Těmto několika silně napojeným uzlům se říká huby. Hub může být např. účet s velkým počtem sledujících v sociální síti, nebo hlavní nádraží v síti železniční. [15]
- Small world — což lze ilustrovat tvrzením, že s mými sousedy máme společné sousedy. Formálně tato vlastnost znamená, že uzly v síti tvoří klastry a mezi dvojicemi uzlů jsou malé vzdálenosti. Příkladem opět může být fenomén sociálních sítí — seznam společných přátel. [16]

V naší práci reprezentujeme síť grafem, tedy množinu uzlů reprezentuje množina vrcholů a vazby jsou převedeny na hrany. Graf triviálně zachovává všechny strukturální vlastnosti sítě, kterou reprezentuje, výše i níže uvedené vlastnosti tedy lze definovat jak pro síť, tak pro graf.

Definice 13 (hustota sítě). *Mějme síť S s n uzly a m vazbami. Hustotu sítě definujeme jako:*

$$\rho = \frac{m}{\binom{n}{2}}.$$

Hustota sítě vlastně vyjadřuje jaké procento z maximálního možného počtu vazeb tvoří skutečný počet vazeb.

Definice 14 (průměrná nejkratší cesta). *Průměrná nejkratší cesta v souvislém grafu G je průměr délek nejkratších cest mezi všemi dvojicemi vrcholů grafu G :*

$$L = \frac{1}{n(n-1)} \sum_{u \neq v} d(u, v)$$

kde $d(u, v)$ značí nejkratší cestu mezi vrcholy u, v .

Definice 15 (lokální klastrovací koeficient). *Nechť v je vrchol v grafu G . Jako lokální klastrovací koeficient vrcholu v značíme poměr počtu hran vedoucích mezi jeho sousedy k počtu možných párů jeho sousedů:*

$$C(v) = \frac{2|E(G(N(v)))|}{\deg(v)(\deg(v) - 1)}$$

Jedním z možných způsobů jak klastrovací koeficient vnímat je jako pravděpodobnost, že jsou sousedi vrcholu v sousedy.

Definice 16 (klastrovací koeficient). *Jako klastrovací koeficient grafu G s n vrcholy označujeme průměr klastrovacích koeficientů všech vrcholů G :*

$$C(G) = \frac{1}{n} \sum_{v \in G} C(v)$$

1.4 Korelace a vzájemná informace

Tato podkapitola čerpá z kurzu Úvod do strojového učení v Pythonu vyučovaném na MFF (NPFL129).

Definice 17 (střední hodnota). *Nechť X je diskrétní náhodná veličina. Jako střední hodnotu x označíme:*

$$E[X] = \sum_{x \in X} x p(x)$$

kde $p(x)$ je zkratka pro $P(X = x)$.

Definice 18 (rozptyl). *Pro náhodnou veličinu X definujeme její rozptyl jako:*

$$\text{var}(X) = E((X - E(X))^2)$$

Definice 19 (kovariance). *Nechť X a Y jsou náhodné veličiny. Jako jejich kovarianci označíme:*

$$\text{cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

Následující veličiny jsme v naší práci využili pro ohodnocení vazeb sítě akcií — za náhodné veličiny zmíněné v definicích jsme dosadili ceny akcií.

Definice 20 (Pearsonův korelační koeficient). *Nechť x a y jsou náhodné veličiny. Jejich Pearsonův korelační koeficient je definován následovně:*

$$r(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}}$$

Definice 21 (vzájemná informace). *Nechť X a Y jsou náhodné veličiny. Řekneme, že jejich vzájemná informace je:*

$$I(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x) p(y)}$$

Mějme 2 náhodné veličiny X a Y . Vzájemná informace $I(X, Y)$ říká, kolik informace tyto 2 veličiny sdílejí, tedy kolik se dozvíme o X , když známe Y a naopak. Jsou-li X a Y nezávislé, pak se ze znalosti X nedozvíme o Y nic a tedy $I(X, Y) = 0$. Naopak, je-li x funkcí y , dostaneme ze znalosti y plnou znalost x a tedy $I(X, Y) = 1$.

2 Stock network

Stock Network je tvořena na základě vývoje cen jednotlivých akcií nějakého konkrétního indexu v daném časovém okně [9]. Uzly jsou tituly (akcie) a vazby vyjadřují podobnost vývoje jejich cen.

2.1 Data

Nejprve potřebujeme vybrat množinu dat, kterou budeme sledovat — v našem případě burzovní index. Vybrali jsme si indexy S&P 500, S&P 600 a RUSSELL 1000 jakožto zástupce amerického trhu — konkrétně burz New York Stock Exchange (zkráceně NYSE) a National Association of Securities Dealers Automated Quotations (zkráceně NASDAQ) — a indexy FTSE 100, FTSE 250 jakožto zástupce trhu evropského — London Stock Exchange (zkráceně LSE). Součástí názvů těchto indexů jsou i počty akcií, které obsahují. Index S&P 500 obsahuje 500 firem s největším obchodovaným objemem na americkém trhu s cennými papíry. RUSSELL 1000 je postaven obdobně, firem ale obsahuje 1000.

Index S&P 600 obsahuje 600 firem s obchodovaným objemem zhruba mezi 250 miliony a 2 miliardami Amerických dolarů. Slovíčko zhruba je v předchozí větě použito, protože neplatí, že jakmile se obchodovaný objem firmy dostane mimo uvedené rozpětí, je firma z indexu vyřazena.

FTSE 100 obsahuje 100 firem s největším obchodovaným objemem na Londýnské burze, FTSE 250 obsahuje 250 firem, které by se v pomyslném žebříčku umístily za firmami indexu FTSE 100. Oba tyto indexy — spolu s indexem RUSSELL 1000 — jsou spravovány společností FTSE Group. S&P 500 spravuje S&P Dow Jones Indices, zatímco S&P 600 spravuje S&P Global Ratings. Kromě toho, že jde o jedny z nejpobulárnějších indexů na světě, byly tyto indexy vybrány také z důvodu snadného získání informací o jejich aktuálním složení [17, 18, 19, 20, 21].

2.2 Parametry dat, jejich získání a zpracování

Nyní si musíme určit období, které budeme zkoumat. Pro naši analýzu jsme si vybrali období od začátku roku 2004 do začátku roku 2024. Tedy obsahující jak finanční krizi let 2007 a 2008, tak uzavření světové ekonomiky kvůli pandemii nemoci Covid-19 v březnu 2020 a rok vypuknutí války na Ukrajině 2022. V tomto období se obchodovalo — a tedy obsahuje data z — 5033 dnů na americkém akciovém trhu a 5051 na London Stock Exchange.

Co se týče samotné hodnoty, kterou sledujeme, vybrali jsme tzv. upravenou uzavřenou cenu — budeme používat běžně užívanou anglickou zkratku *Adj. Close*. Tato hodnota započítává — na rozdíl od běžné uzavřené ceny, což je cena posledního obchodu provedeného před uzavřením trhu — vše, co může cenu ovlivnit po uzavření trhu. To může být např. štěpení akcií, vyplácení dividend nebo přidělení výhradních práv. Z těchto důvodů se *Adj. Close* běžně používá při zkoumání historického vývoje akcií.

Tyto data stahujeme pomocí API z Yahoo Finance [22]. Po načtení data zkontrolujeme na možné výpadky, popř. chybějící data dopočteme lineární interpolací.

Na samotnou analýzu si ponecháme pouze takové akcie, které jsou definované po celou dobu časového období, které zkoumáme. Zde čtenáře možná překvapí, že skutečně použitých akcií indexu S&P 600 je méně, než indexu S&P 500 (tedy na počet akcií menšího indexu). Jedním z možných vysvětlení je, že zatímco S&P 600 obsahuje akcie firem s malou tržní kapitalizací (jedná se o tzv. small cap index), S&P 500 obsahuje akcie firem s velkou tržní kapitalizací (je tzv. large cap) a tedy firem, které jsou na trhu typicky déle (viz tabulka 2.1). Všechny parametry dat jsou shrnuty v tabulce 2.2.

Jakmile máme načtená data, která jsou navíc zkontrolována na výpadky, spočteme z nich logaritmus a to podle následujícího vzorce:

$$\log(r_i) - \log(r_{i+1})$$

kde r_i je cena akcie v i -tý den časového období.

2.3 Parametry sítě a sestavení grafu sítě

Naši síť stavíme vždy pro časové okno, potřebujeme tedy určit, jak bude časové okno dlouhé a jak moc jej budeme posouvat. Pro naši analýzu jsme zvolili časová okna o délce 3, 6 a 12 měsíců a posouvali jsme je vždy o 1 měsíc — to znamená, že začneme-li analýzu v lednu 2004, pak první okno délky 3 měsíce obsahuje leden, únor a březen, druhé okno únor, březen a duben, atd. Pro všechny zkoumané burzy každé okno délky 3 měsíce obsahuje od 59 do 66 dnů, ze kterých máme data. U oken délky 6 měsíců toto činí 121 až 132 dnů a u oken délky 12 měsíců 248 až 261 dnů.

Pro každé časové okno postavíme graf sítě tak, že se podíváme na podobnost vývoje cen všech dvojic akcií. Pro měření podobnosti jsme vybrali korelační koeficient v absolutní hodnotě a společnou informaci. Každá dvojice vrcholů — akcií — je spojena hranou ohodnocenou podobností vývoje jejich ceny v čase. Společnou informaci jsme odhadovali algoritmem EQQ[23], který jako argument bere počet binů. Do něj jsme dosazovali 4, 6 a 8 pro časová okna velikosti 3 a 6 měsíců a pro velikost 12 měsíců jsme dosadili 6 a 8. Více o algoritmu EQQ a možných volbách parametru binů lze dočíst v článku Hartmana a Hlinky [24]. Všechny parametry sestavování sítě a jejího grafu jsou opět uvedeny v tabulce 2.2.

Tímto postupem dostaneme úplný graf, protože podobnost vývoje dvou akcií nikdy není přesně nulová. Úplný graf je ale úplně automorfnní — všechny vrcholy sousedí se všemi a tedy by nám měření symetrie nic neřeklo.

Index	Počet použitých akcií
S&P 500	391
S&P 600	348
RUSSELL 1000	612
FTSE 100	81
FTSE 250	134

Tabulka 2.1 Skutečně použité počty akcií.

V úvodu jsme zmínili, proč pro zkoumání symetrií nedává smysl stavění maximální kostry a také jsme zmínili 2 další přístupy.

Maximální rovinný filtrovaný graf je rovinný graf postupně skládaný z hran podle jejich váhy — to je od hran s největší vahou po ty nejmenší — s pravidlem, že když by přidání hrany porušilo rovinnost grafu, je tato hrana zkrátka přeskočena a do grafu se nepřidá. Hrany jsou do grafu přidávány tak dlouho, dokud není jejich počet v něm maximální možný pro rovinný graf dané velikosti [11]. Takto získaný graf tedy nedodrží pořadí hran dané jejich ohodnocením a proto nám nevyhovuje.

Strategie vítěz bere vše je parametrizována číslem t , a do výsledného grafu zahrne všechny hrany s vahou alespoň t [10]. Takto získaný graf ale může být — a v našem případě často je — nesouvislý. To nám nevyhovuje, protože budou-li se v grafu vyskytovat izolované vrcholy, mohou se na sebe zobrazovat automorfismem a tím významně ovlivnit velikost automorfnní grupy. Toto neplatí jen pro izolované vrcholy, ale také pro menší komponenty, které se mohou vyskytovat úplně.

Ani jedna ze 3 zmíněných metod k získání grafu sítě nám tedy nevyhovuje. Navrhli jsme proto 3 metody, které v následující podkapitole postupně představíme.

Parametr	Hodnoty
Zdroj dat	Yahoo Finance [22]
Časové období	2004-01-01 až 2024-01-01
Burza	NYSE, NASDAQ, LSE
Burzovní idnex	S&P 500, S&P 600, RUSSELL 1000, FTSE 100, FTSE 250,
Typ hodnoty	Adj. Close
Časové okno	3 měsíce, 59 až 66 dnů 6 měsíců, 121 až 132 dnů 12 měsíců, 248 až 261 dnů
Posun časového okna	1 měsíc
Vyjádření podobnosti vývoje ceny	korelace, vzájemná informace
Počet binů pro vzájemnou informaci	4, 6, 8 pro časová okna 3 a 6 6 a 8 pro časové okno 12
Cílovaná hustota sítě	0,2

Tabulka 2.2 Tabulka všech parametrů.

2.4 Získávání komponenty souvislosti

Ve výzkumu se běžně používají sítě s nízkou hustotou [7] [10]. Také bychom proto chtěli síť s nízkou hustotou, zároveň ale chceme, aby byla síť souvislá a skládala se z co nejvíce uzlů. Dále chceme, aby parametry grafu sítě — konkrétně hustota a velikost — byly napříč časovými okny co nejpodobnější. Rozhodli jsme se proto získat největší komponentu souvislosti a cílit různým způsobem její hustotu na hodnotu 0,2 a velikost na co nejvíce podobnou velikosti celkové sítě. Toho jsme se snažili dosáhnout třemi metodami.

Jako první představíme metodu 1, ve které postupně odebíráme hrany seřazené podle jejich vah a u toho se díváme na největší komponentu s její hustotou. Hrany odebíráme tak dlouho, dokud největší komponenta nemá požadovanou hustotu, nebo dokud nám nedojdou hrany (viz algoritmus 1).

Hlavní výhodou této metody je stálá hustota komponenty napříč časovými okny.

Nevýhodou je, že ne vždy lze postupným odebíráním hran komponentu požadované hustoty získat. To se pak odráží v její velikosti, jelikož na tu při tomto postupu vůbec nehledíme. V krajních případech — v některých časových oknech pro FTSE 100 (viz obrázek 2.1) a FTSE 250 — skončíme s komponentou velikosti 1, na které pak nic nenaměříme. I mimo tyto případy je však přílišná fluktuace velikosti komponenty nežádoucí, protože symetričnost sítě se poměrně obtížně normuje — každý vrchol navíc totiž může být předmětem stávajících automorfismů, může zapříčinit automorfismy nové, nebo nemusí udělat ani jedno a stávající automorfni grupu tak nezměnit vůbec.

Algoritmus 1 První metoda získávání komponenty souvislosti. Při odebírání hran sledujeme největší komponentu souvislosti a zároveň s ní její hustotu. Hrany odebíráme tak dlouho, dokud nemáme souvislý graf s požadovanou hustotou (s danou tolerancí c), nebo dokud nám nedojdou.

```
1: function FITDENSITY1( $G, \rho, c$ )
2:    $E \leftarrow$  hrany grafu  $G$  seřazené podle váhy
3:    $H \leftarrow$  Giant( $G$ )
4:    $\rho_H \leftarrow$  hustota  $H$ 
5:    $i \leftarrow 0$ 
6:   while  $i < \text{len}(E)$  &  $\rho_H > \rho + c$  do
7:      $e \leftarrow E[i]$ 
8:      $G \leftarrow G \setminus e$ 
9:     if  $e \in H$  then
10:       $H \leftarrow H \setminus e$ 
11:      if  $H$  není souvislý then
12:         $H \leftarrow$  Giant( $G$ )
13:      end if
14:       $\rho_H \leftarrow$  hustota  $H$ 
15:    end if
16:     $i \leftarrow i + 1$ 
17:  end while
18:  Vrátíme  $G, H, \rho_H$ 
19: end function
```

Ve druhé metodě postupujeme poněkud příměji. Odebíráním hran postupně získáváme graf s hustotou stále bližší té požadované. Jakmile se dostaneme do hodnoty požadovaná hustota + c , začneme měřit velikost největší souvislé komponenty a pokračujeme v odebírání hran, dokud velikost komponenty neklesne, nebo dokud se nedostaneme k požadované hodnotě hustoty sítě (viz algoritmus 2).

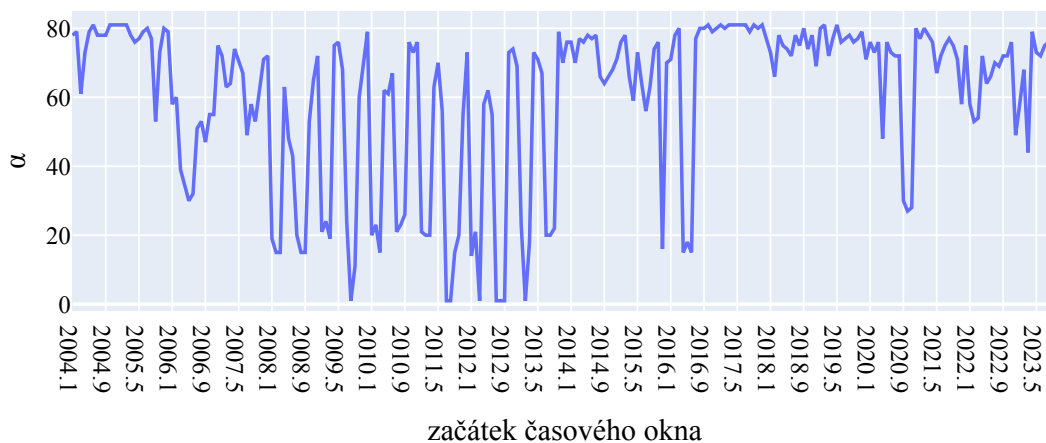
Oproti minulému postupu už nám nedělají extrémní případy takový problém a pro rozumně nastavené c (zvolili jsme 0,05) dostáváme grafy s relativně stálým počtem vrcholů i hustotou. Na druhou stranu nám stále dělají největší problém konfigurace s indexy FTSE 100 (viz obrázek 2.2) a FTSE 250, ohodnocené koralčím koeficientem a s velikostí okna 3 měsíce, kde se v některých časových oknech dostáváme na téměř polovinu vrcholů.

Algoritmus 2 Druhá metoda získávání komponenty souvislosti. Odebíráním hran získáme graf s maximální přijatelnou hustotou, poté sledujeme velikost největší komponenty a dalším odebíráním hran se přibližujeme k požadované hustotě, dokud se nám největší komponenta nezmenší, nebo dokud se k požadované hustotě nedostaneme.

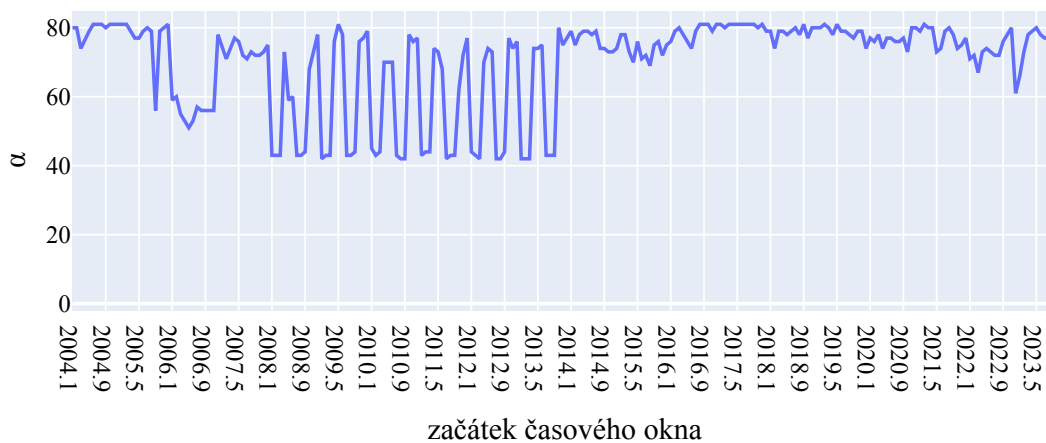
```

1: function FITDENSITY2( $G, \rho, c$ )
2:    $E \leftarrow$  hrany grafu  $G$  seřazené podle váhy
3:    $\rho_G \leftarrow$  hustota  $G$ 
4:    $i \leftarrow 0$ 
5:   while  $i < \text{len}(E)$  &  $\rho_G > \rho + c$  do
6:      $e \leftarrow E[i]$ 
7:      $G \leftarrow G \setminus e$ 
8:      $\rho_G \leftarrow$  hustota  $G$ 
9:      $i \leftarrow i + 1$ 
10:  end while
11:   $H \leftarrow \text{Giant}(G)$ 
12:   $V_H \leftarrow$  vrcholy  $H$ 
13:   $V'_H \leftarrow V_H$ 
14:  while  $i < \text{len}(E)$  &  $\rho_G > \rho$  &  $\text{len}(V_H) == \text{len}(V'_H)$  do
15:     $e \leftarrow E[i]$ 
16:     $G \leftarrow G \setminus e$ 
17:    if  $e \in H$  then
18:       $H \leftarrow H \setminus e$ 
19:      if  $H$  není souvislý then
20:         $V_H \leftarrow$  vrcholy  $\text{Giant}(G)$ 
21:        if  $\text{len}(V_H) == \text{len}(V'_H)$  then
22:           $H \leftarrow \text{Giant}(G)$ 
23:        end if
24:      end if
25:    end if
26:     $\rho_G \leftarrow$  hustota  $G$ 
27:     $i \leftarrow i + 1$ 
28:  end while
29:   $H \leftarrow H + e$ 
30:   $\rho_H \leftarrow$  hustota  $H$ 
31:  Vrátíme  $G, H, \rho_H$ 
32: end function

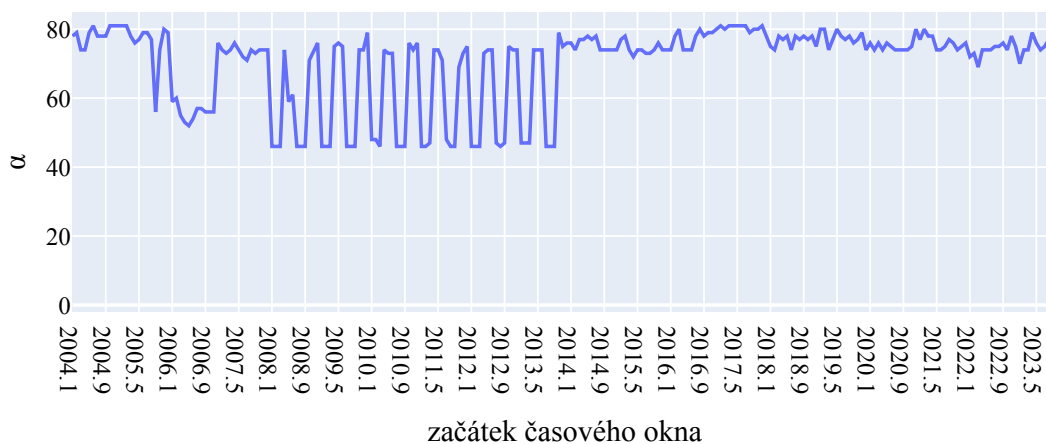
```



Obrázek 2.1 Počet vrcholů největší komponenty souvislosti grafu sítě FTSE 100 ohodnocené korelací v časových oknech velikosti 3 měsíce vytvořené metodou 1.



Obrázek 2.2 Počet vrcholů největší komponenty souvislosti grafu sítě FTSE 100 ohodnocené korelací v časových oknech velikosti 3 měsíce vytvořené metodou 2.



Obrázek 2.3 Počet vrcholů největší komponenty souvislosti grafu sítě FTSE 100 ohodnocené korelací v časových oknech velikosti 3 měsíce vytvořené metodou 3.

Třetí metoda je vlastně zobecněním druhé, nebudeme proto uvádět pseudokód, ale pouze toto zobecnění popíšeme. Místo podmínky $\text{len}(V_H) == \text{len}(V'_H)$ na 14. a 21. řádku řekneme, že počet vrcholů největší komponenty nesmí být menší, než nějaký t -násobek počtu vrcholů celého grafu: $\text{len}(V_H) \geq t * n$. My jsme zvolili $t = 0,84$, čímž dostáváme podle nás nejmenší dostatečně velkou komponentu. Zároveň u této metody můžeme rozvolnit parametr c — my jsme jej navýšili na 0,1 — protože už velikost největší komponenty není dána prvním vstupem do while cyklu na řádku 14. Tímto jsme velikosti největší komponenty souvislosti u nejhorsích případů zploštili (viz obrázek 2.3).

2.5 Měřené veličiny

Ještě než začneme se samotnou analýzou, musíme si určit, co přesně budeme na jednotlivých sítích měřit. Symetrii, budeme měřit — tak jako MacArthur [7] — pomocí automorfismů.

Definice 22 (velikost automorfní grupy). *Jako velikost automorfní grupy grafu G bereme počet jejích prvků, tedy počet automorfismů, které na grafu existují:*

$$\alpha = |\text{Aut}(G)|$$

Tato veličina je silně ovlivněna velikostí grafu, pracovali jsme proto ještě s normovaným počtem orbit.

Definice 23 (normovaný počet orbit). *Nechť n je počet vrcholů grafu G a $O(G)$ je množina všech orbit všech automorfismů na grafu G . Pak jako r označíme normovaný počet orbit:*

$$r = \frac{|O(G)| - 1}{n - 1}$$

Tato definice je přezvata z článku Balla a Schulze [13]. Původně normovaný počet orbit zavedl MacArthur pod názvem *network redundancy* [7] a to ve tvaru

$$r = \frac{|O(G)| - 1}{n}$$

Ve svém článku pomocí něj vyhodnocoval zranitelnost sítě vůči útokům. Prvky v rámci jedné orbity si navzájem slouží jako zálohy, kdyby byl jeden z nich napaden [7]. My jsme zvolili verzi Balla a Schulze, protože pokud je G asymetrický — tedy jediným prvkem $\text{Aut}(G)$ je identita — pak všechny orbity jsou velikosti 1, $|O(G)| = n$ a tedy $r = 1$, což v původní MacArthurově verzi neplatí.

Měřili jsme také nenormovaný počet orbit, a to jak všech, tak těch netriviálních (tedy s více než jedním prvkem) a dívali jsme se na distribuci velikostí a složení orbit.

2.6 Obarvení grafu podle GICS

Měření výše uvedených veličin jsme prováděli také na obarveném grafu. To znamená, že automorfismus na sebe může zobrazovat vrcholy pouze stejné barvy. Vrcholy jsme obarvili podle jejich GICS zařazení [25].

GICS — zkratka pro Global Industry Classification Standard — rozděluje firmy podle jejich působení do 11 sektorů. Zatímco americké indexy mají informaci o svém GICS sektoru přímo, evropské (FTSE 100 a FTSE 250) používají jiný systém klasifikace — konkrétně ICB, tedy Industry Classification Benchmark, který firmy rozděluje do 11 odvětví. Toto jsme vyřešili jednoduchým převedením, protože 11 sektorů GICS zhruba odpovídá 11 odvětvím ICB. Za zmínku stojí, že omezením na vrcholy stejné barvy většinou moc automorfismů nebylo. To dává smysl. Těžařská firma se bude spíš chovat jako jiná těžařská firma, než jako firma provozující síť restaurací. Počty skutečně použitých akcií v jednotlivých sektorech jsou uvedeny v tabulce 2.3.

2.7 Použitý software

Na závěr této kapitoly zmííme nástroje, které jsme v naší práci použili. K načítání dat a samotné analýze jsme používali balíček, který vznikl v rámci ročníkového projektu [26] (s rozšířením o metody získávání největší komponenty souvislosti). Balíček využívá několik knihoven — za zmínku stojí hlavně ty z prostředí sage [27] — a umožňuje jednoduše parametrizovat jak požadovaná data, tak síť, kterou chceme analyzovat.

Data načítáme z Yahoo Finance [22] prostřednictvím pythonské knihovny pandas-datareader [28]. K výpočtu vzájemné informace jsme použili funkci `mual_information` z balíčku `pyCliTS` [23].

Index	1	2	3	4	5	6	7	8	9	10	11
S&P 500	18	22	58	39	31	55	57	47	10	28	26
S&P 600	12	21	67	57	22	28	65	41	6	10	19
RUSSELL 1000	26	41	105	73	39	76	94	66	15	35	42
FTSE 100	2	4	19	19	7	3	19	2	1	4	1
FTSE 250	3	4	21	22	4	1	71	6	2	0	0

Tabulka 2.3 Počty skutečně použitých akcií podle industriálních sektorů GICS. 1–Energy, 2–Materials, 3–Industrials, 4–Consumer Discretionary, 5–Consumer Staples, 6–Health Care, 7–Financials, 8–Information Technology, 9–Communication Services, 10–Utilities, 11–Real Estate

3 Analýza sítí

Nyní máme vše připraveno a můžeme přistoupit k samotné analýze. Celkem jsme vyhodnotili 33 konfigurací pro každý index. Konkrétně jsme vyhodnotili síť ohodnocené korelačním koeficientem vytvořené všemi 3 metodami, pro všechny 3 velikosti časového okna. Celkově tedy 9 konfigurací s korelačním koeficientem pro každý index.

Sítě ohodnocené vzájemnou informací jsme taktéž vyhodnotili všemi metodami. U časových oken velikost 3 a 6 jsme dosazovali parametr binů 3, 6 a 8, u časového okna velikosti 12 jsme dosazovali 6 a 8 binů. Celkově tedy 24 konfigurací s ohodnocením sítě vzájemnou informací pro každý index.

Níže uvádíme příklady, které vyšly nejlépe — tedy hodnoty α a r nabývají viditelně odlišných hodnot v časových oknech obsahující propady na světovém trhu. Konkrétně těmito propady myslíme světovou finanční krizi let 2007 a 2008, vypuknutí pandemie nemoci Covid-19 v březnu 2020 a první rok války na Ukrajině 2022.

Příklady budeme procházet v pořadí podle použitých metod (popsaných v sekci 2.4) získání největší komponenty souvislosti. U každé metody ukážeme alespoň 1 příklad sítě ohodnocené korelačním koeficientem a alespoň 1 příklad sítě ohodnocené vzájemnou informací. Uváděná hustota je vždy zaokrouhlena na tisíce. U orbit, v nichž všechny prvky patří do stejného sektoru GICS tento sektor uvedeme.

3.1 Metoda 1

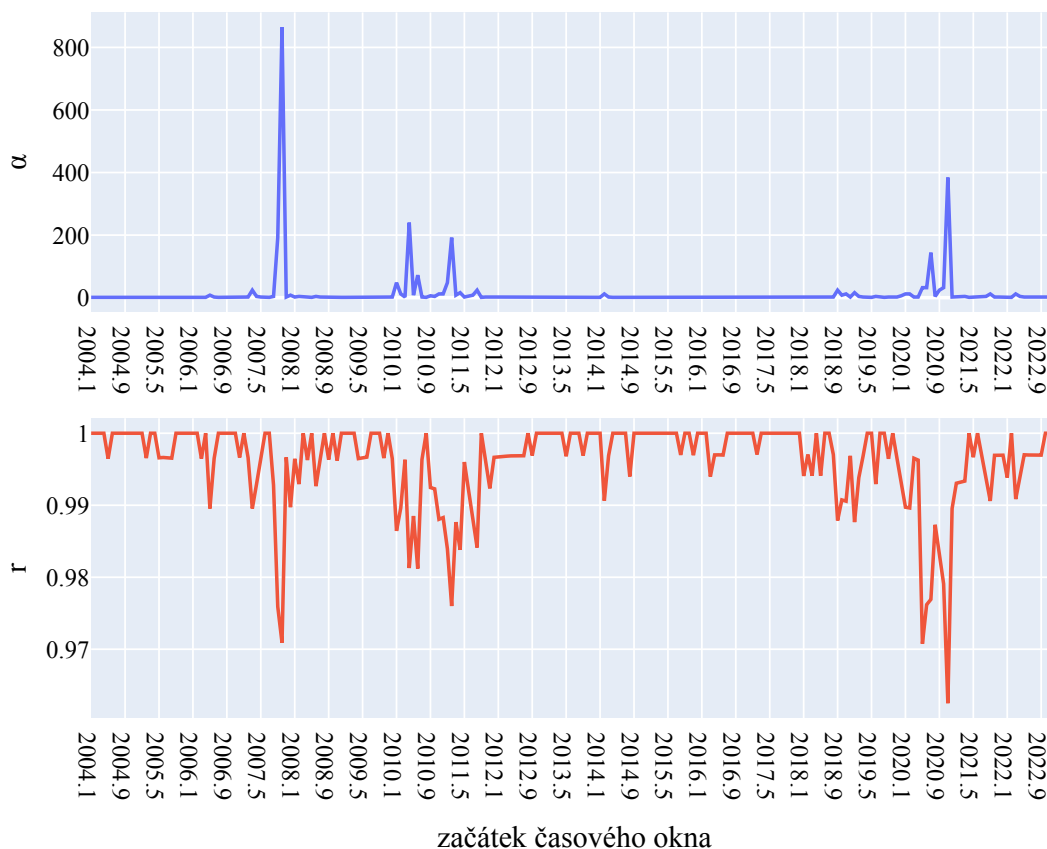
Metoda 1 cílila hodnotu hustoty, která po zaokrouhlení na tisíce vždy vyšla stejně — tedy cílená hustota $0,2 + \text{tolerance } c = 0,05 - 0,25$. U procházených příkladů ji tedy neuvádíme.

U sítí ohodnocených korelačním koeficientem vyšla nejlépe konfigurace s indexem S&P 600 při velikosti časového okna 12 měsíců (viz obrázek 3.1).

Nejvíce vystupuje okno zahrnující světovou finanční krizi (1. 10. 2007 až 1. 9. 2008) kde je velikost automorfí grupy rovna 864 automorfismů. Automorfismy zde tvoří 3 orbity velikosti 3: (CLB, HP, OII) – Energy, (HLX, NBR, PTEN) – Energy, (MSEX, RDNT, SANM). Dále 2 orbity velikosti 2: (HLIT, PRG), (ATGE, STRA) – Consumer Discretionary. Zbytek už jsou orbity triviální — tedy velikosti 1 — konkrétně je jich 263. Síť tedy obsahuje 276 uzlů.

Dále vystupuje okno začínající v podzimním znovuuzavírání ekonomik kvůli pandemii nemoci Covid-19 (1. 11. 2020 až 1. 10. 2021). Automorfí grupa v něm obsahuje 384 automorfismů, které tvoří 1 orbitu velikosti 3 (HTLD, HUBG, MRTN) – Industrials, 7 orbit velikosti 2: (AEIS, VECO) – Information Technology, (AVA, SJW) – Utilities, (AWR, NWN) – Utilities, (CEVA, COHU) – Information Technology, (HVT, UPBD) – Consumer Discretionary, (MAC, SKT) – Real Estate, (NBR, RES) – Energy a 224 orbit velikosti 1. Síť má 241 uzlů.

V pořadí další vystupující okno (1. 4. 2010 až 1. 3. 2011) s velikostí automorfí grupy 240 již zajímavé vychýlení nevykazuje — nezasahuje ani do jedné z námi sledovaných událostí. Podívejme se proto na stejnou síť, kde automorfismy budou respektovat obarvení podle GICS (viz obrázek 3.2).



Obrázek 3.1 Velikost automorfnní grupy α a normovaný počet orbit r největší komponenty souvislosti grafu sítě S&P 600 ohodnocené korelačním koeficientem v časových oknech velikosti 12 měsíců vytvořené metodou 1.

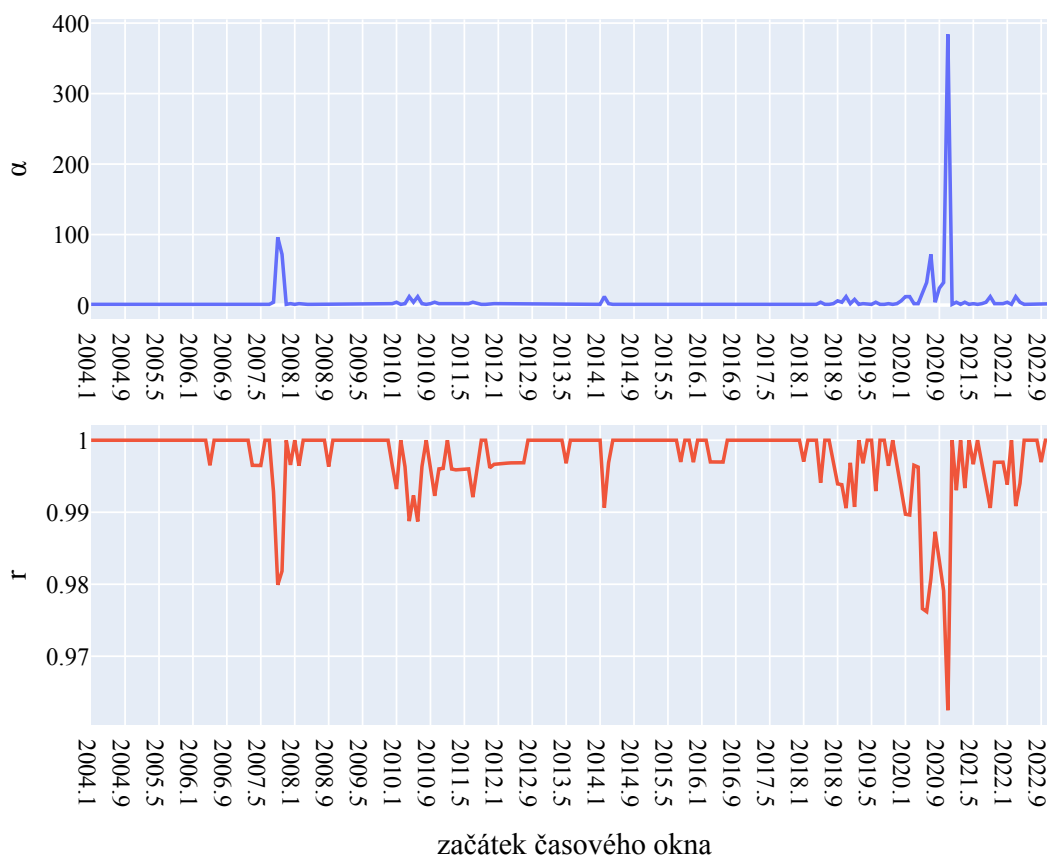
Zde nejvíce vystupuje poslední zmíněné okno (1. 11. 2020 až 1. 10. 2021), jehož vlastnosti se vůbec nezměnily — všechny automorfismy zde existovaly už v rámci stejných GICS sektorů.

Další vystupuje okno (1. 9. 2007 až 1. 8. 2008) s velikostí automorfnní grupy 96 automorfismů. Ty v tomto okně tvoří 1 orbitu velikosti 4: (CLB, DRQ, HLX, SM) – Energy, 2 orbity velikosti 2: (HP, NBR) – Energy, (ATGE, STRA) – Consumer Discretionary a 242 orbit velikosti 1. Síť má 250 uzlů.

Jako třetí vystupuje opět okno (1. 10. 2007 až 1. 9. 2008), tentokrát však s velikostí automorfnní grupy 72. Ta se tedy výrazně zmenšila, z orbit se zachovaly 2 velikosti 3: (CLB, HP, OII) – Energy, (HLX, NBR, PTEN) – Energy, 1 velikosti 2: (ATGE, STRA) – Consumer Discretionary a tím pádem se zvětšil počet orbit velikosti 1 na 268. Na velikost sítě obarvení nemá vliv a tento parametr tak zůstal stejný — 276 uzlů.

Se stejnou velikostí automorfnní grupy se ještě objevuje okno obsahující opět znovuzavření ekonomik na podzim roku 2020 (1. 7. 2020 až 1. 6. 2021). To samé platí pro stukturu automorfismů, které tvoří 2 orbity velikosti 3: (NBR, RES, SM) – Energy, (CVCO, MDC, MHO) – Consumer Discretionary, 1 orbitu velikosti 2: (ETD, HVT) – Consumer Discretionary a 251 orbit velikosti 1. Síť má tedy 261 uzlů.

Ze sítí ohodnocených vzájemnou informací vyšla nejlépe konfigurace s indexem



Obrázek 3.2 Velikost automorfnní grupy α a normovaný počet orbit r respektující obarvení podle GICS největší komponenty souvislosti grafu sítě S&P 600 ohodnocené korelačním koeficientem v časových oknech velikosti 12 měsíců vytvořené metodou 1.

S&P 600 při velikosti časového okna 6 měsíců a počtem binů 4 (viz obrázek 3.3). V této konfiguraci nejvíce vystupuje okno zahrnující světovou finanční krizi (1. 3. 2008 až 1. 8. 2008) s velikostí automorfnní grupy 96 automorfismů. Ty zde tvoří 1 orbitu velikosti 4: (CLB, DRQ, HLX, SM) – Energy, 2 orbity velikosti 2: (CRK, HP) – Energy, (SNEX, TDS) a 311 orbit velikosti 1. Síť tedy obsahuje 319 uzlů.

Další je časové okno zahrnující celosvětové uzavření ekonomik kvůli pandemii nemoci Covid-19 (1. 3. 2020 až 1. 8. 2020) s velikostí automorfnní grupy 48 automorfismů. Zde automorfismy tvoří 1 orbitu velikosti 3: (CWT, MGEE, SJW) – Utilities, 3 orbity velikosti 2: (AVA, AWR) – Utilities, (ETD, HVT) – Consumer Discretionary, (HZO, NUS) a 283 orbit velikosti 1. Síť je tedy tvořena 292 uzly.

V pořadí třetí časové okno zahrnuje vypuknutí války na Ukrajině (1. 1. 2022 až 1. 6. 2022) s velikostí automorfnní grupy 24 automorfismů, které tvoří 1 orbitu velikosti 3: (DRQ, HLX, PTEN) – Energy, 2 orbity velikosti 2: (AVA, NWN) – Utilities, (CLB, RES) – Energy a 338 orbit velikosti 1. Síť má 345 uzlů.

Jak u velikosti automorfnní grupy, tak u normovaného počtu orbit si lze všimnout, že v období mezi světovou finanční krizí a pandemií nemoci Covid-19 panuje až na pár drobných výjimek klid.



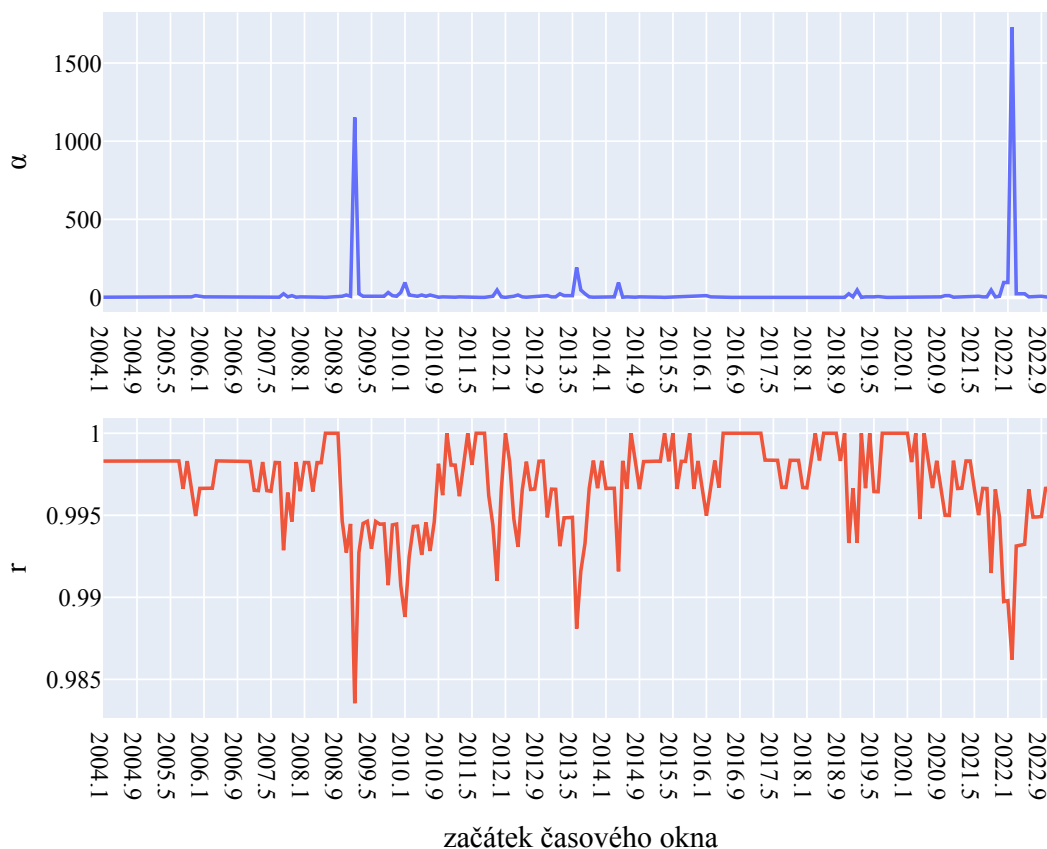
Obrázek 3.3 Velikost automorfnní grupy α a normovaný počet orbit r největší komponenty souvislosti grafu sítě S&P 600 ohodnocené vzájemnou informací s počtem binů 4 v časových oknech velikosti 6 měsíců vytvořené metodou 1.

3.2 Metoda 2

U druhé metody vyšel ze sítě ohodnocených korelačním koeficientem nejlépe index RUSSELL 1000 při velikosti okna 12 měsíců (viz obrázek 3.4). Zde nejvíce vystupuje okno obsahující válku na Ukrajině (1. 2. 2022 až 1. 1. 2023), kde je velikost automorfnní grupy 1728 automorfismů. Ty zde tvoří 1 orbitu velikosti 4: (CPB, CAG, GIS, K) – Consumer Staples, 2 orbity velikosti 3: (CNC, HUM, MOH) – Health Care, (LHX, LMT, NOC) – Industrials, 1 orbitu velikosti 2: (CVX, COP) – Energy a 569 orbit velikosti 1. Síť má velikost 581 uzlů a hustotu 0,274.

Dále vystupuje okno těsně po světové finanční krizi (1. 1. 2009 až 1. 12. 2009), které má velikost automorfnní grupy 1152. Automorfismy zde tvoří 2 orbity velikosti 3: (BMY, MRK, PFE) – Health Care, (GIS, HRL, MKC) – Consumer Staples, 5 orbit velikosti 2: (AAP, AZO) – Consumer Discretionary, (CAH, MCK) – Health Care, (CRL, RMD) – Health Care, (CHE, WTRG), (IAC, MTCH) – Communication Services a 532 orbit velikosti 1. Síť je tvořena 548 uzly a má hustotu 0,318.

V pořadí podle velikosti automorfnní grupy je další okno, které žádný ze sledovaných selhání trhu neobsahuje (1. 6. 2013 až 1. 5. 2014). Podíváme-li se na automorfnní grupu respektující obarvení podle GICS (viz obrázek 3.5), má toto okno automorfnní grupu mnohem menší. Z velikosti 192 se stala 48. Zároveň se při respektování obarvení zmenšila i automorfnní grupa okna těsně po světové



Obrázek 3.4 Velikost automorfnní grupy α a normovaný počet orbit r komponenty souvislosti grafu síte RUSSELL 1000 ohodnocené korelačním koeficientem v časových oknech velikosti 12 měsíců vytvořené metodou 2.

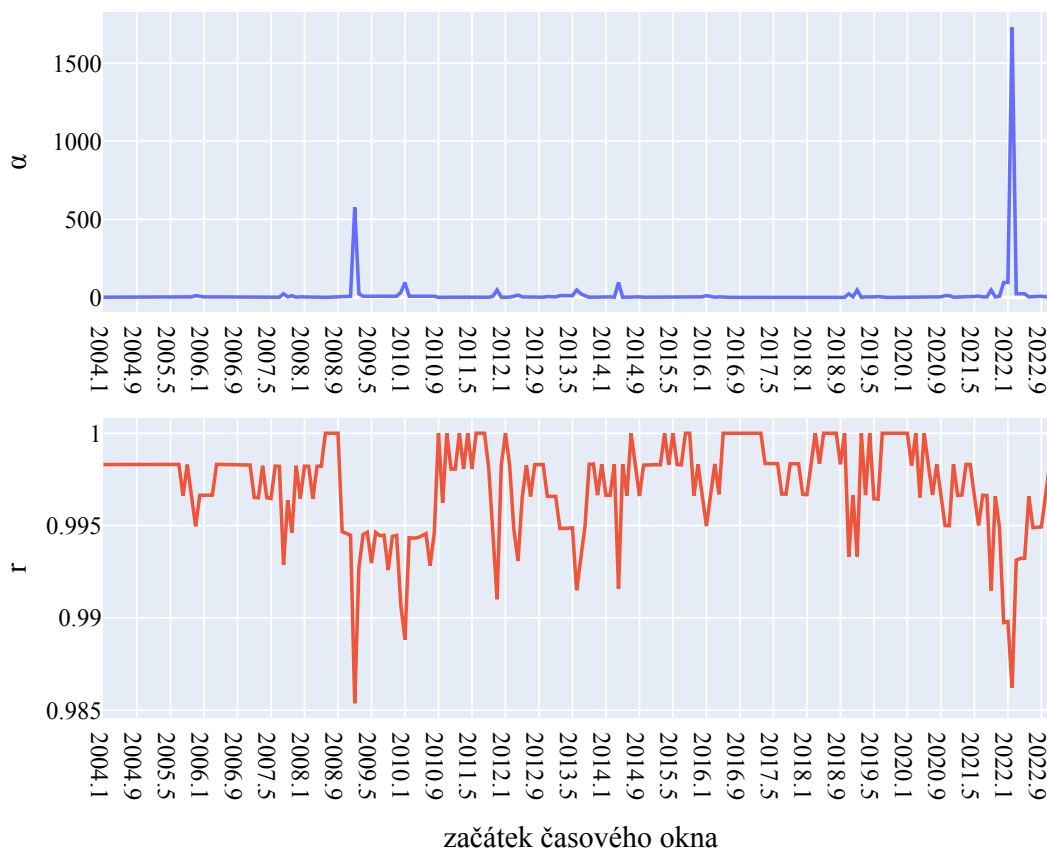
finanční krizi zmíněného vyše (1. 1. 2009 až 1. 12. 2009). Konkrétně z 1152 na 576.

Ze síteí ohodnocených vzájemnou informací vyšel nejlépe index FTSE 100 při velikosti okna 3 měsíce a respektující obarvení podle GICS (viz obrázek 3.6). Odhadu vzájemné informace byl dán argument 6 binů.

Zde jednoznačně vystupují 3 časová okna, každé z nich strefuje přesně události, které chceme. Okno s velikostí automorfnní grupy 4 je těsně po vypuknutí války na ukrajině (1. 3 2022 až 1. 5. 2022) a automorfismy tvoří 2 orbity velikosti 2: (AAL.L, RIO.L) – Materials, (SVT.L, UU.L) – Utilities a 76 orbit velikosti 1. Síť má velikost 80 uzlů a hustotu 0,251.

Další okna (1. 6. 2008 až 1. 8 2008) a (1. 2. 2020 až 1. 4. 2020) obsahují světovou finanční krizi a vypuknutí pandemie nemoci Covid-19. Obě mají velikost automorfnní grupy 2 a obě jsou tím pádem tvořeny 1 orbitou velikosti 2: (AAL.L, ANTO.L) – Materials, resp. (SVT.L, UU.L) – Utilities a 77, resp. 78 orbitami velikosti 1. Uzlů mají síte 79, resp. 80 a jejich hustota je 0,244, resp. 0,277.

Jedinou nevýhodou tohoto příkladu je, že automorfnní grupy jsou dost malé a tedy nelze vyloučit možnost náhodné symetrie. Na druhou stranu by se náhodná symetrie měla náhodně i objevovat a zde máme symetrie přesně ve 3 místech, která pozorujeme. Také se všechny 4 orbity skládají z firem ze stejných sektorů, což jde proti možnosti náhodné symetrie.



Obrázek 3.5 Velikost automorfnní grupy α a normovaný počet orbit r respektující obarvení největší komponenty souvislosti grafu sítě RUSSELL 1000 ohodnocené korelačním koeficientem v časových oknech velikosti 12 měsíců vytvořené metodou 2.

3.3 Metoda 3

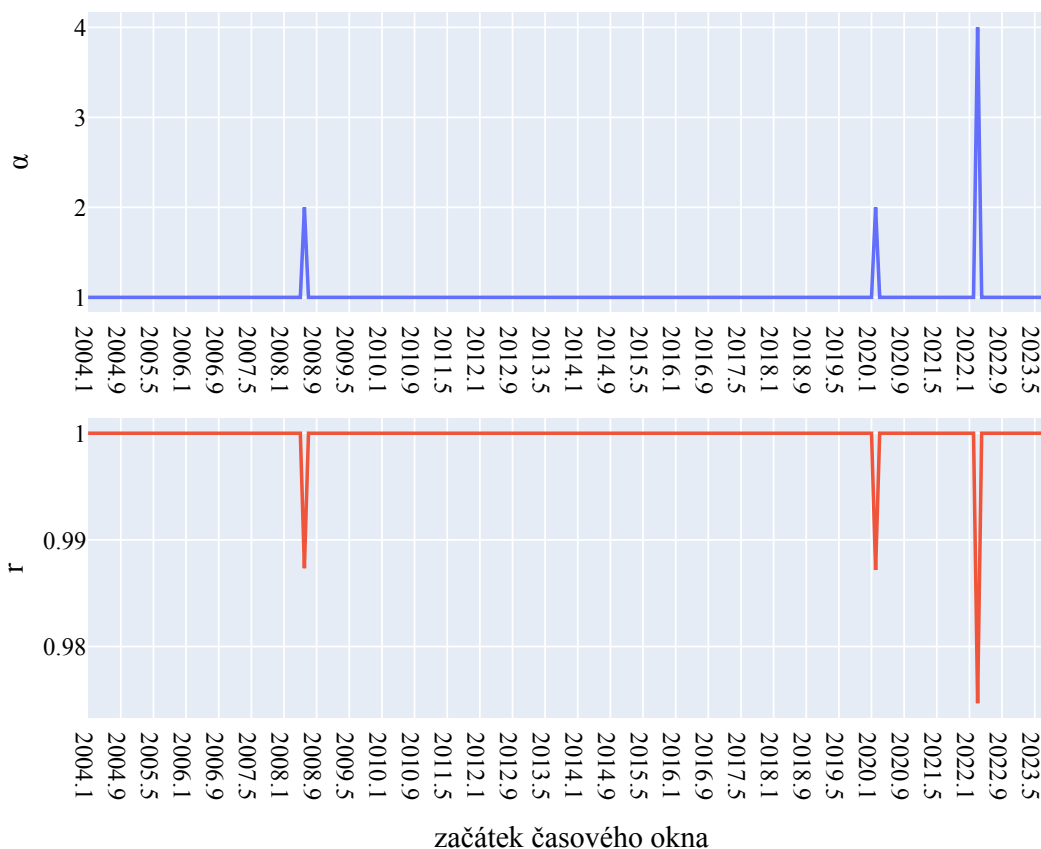
U poslední metody vyšel ze sítí ohodnocených korelačním koeficientem nejlépe index S&P 600 a to při velikosti okna 12 měsíců (viz obrázek 3.7).

Při tomto nastavení viditelně vystupuje okno obsahující světovou finanční krizi (1. 10. 2007 až 1. 9. 2008). Velikost automorfnní grupy je zde 96. Automorfismy v něm tvoří 1 orbitu velikosti 3: (HLX, HP, NBR) – Energy, 4 orbity velikosti 2: (ACLS, NTCT) – Information Technology, (AORT, DDD), (CLB, CRK) – Energy, (CRVL, LGND) – Health Care a 283 orbit velikosti 1. Síť obsahuje 294 uzlů a její hustota je 0,297.

Dále vystupuje okno obsahující válku na Ukrajině (1. 10. 2021 až 1. 9. 2022). Zde je velikost automorfnní grupy rovna 48 a automorfismy tvoří 1 orbitu velikosti 3: (AVA, CPK, MGEE) – Utilities, 3 orbity velikosti 2: (AWR, CWT) – Utilities, (DRQ, WKC) – Energy, (OII, RES) – Energy a 308 orbit velikosti 1. Síť je velká 317 uzlů a její hustota je 0,242.

Co se sítí ohodnocených vzájemnou informací týče, uvádíme opět konfiguraci s indexem S&P 600, tentokrát při velikosti okna 6 měsíců (viz obrázek 3.8). Odhadu vzájemné informace byl dán argument 4 biny.

Viditelně vystupuje okno obsahující válku na Ukrajině (1. 1. 2022 až 1. 6. 2022), velikost automorfnní grupy je zde 72 automorfismů, které tvoří 2 orbity velikosti 3: (AWR, CWT, SJW) – Utilities, (CLB, OII, RES) – Energy, 1 orbitu velikosti 2:



Obrázek 3.6 Velikost automorfnní grupy α a normovaný počet orbit r respektující obarvení největší komponenty souvislosti grafu sítě FTSE 100 ohodnocené vzájemnou informací s počtem binů 6 v časových oknech velikosti 3 měsíce vytvořené metodou 2.

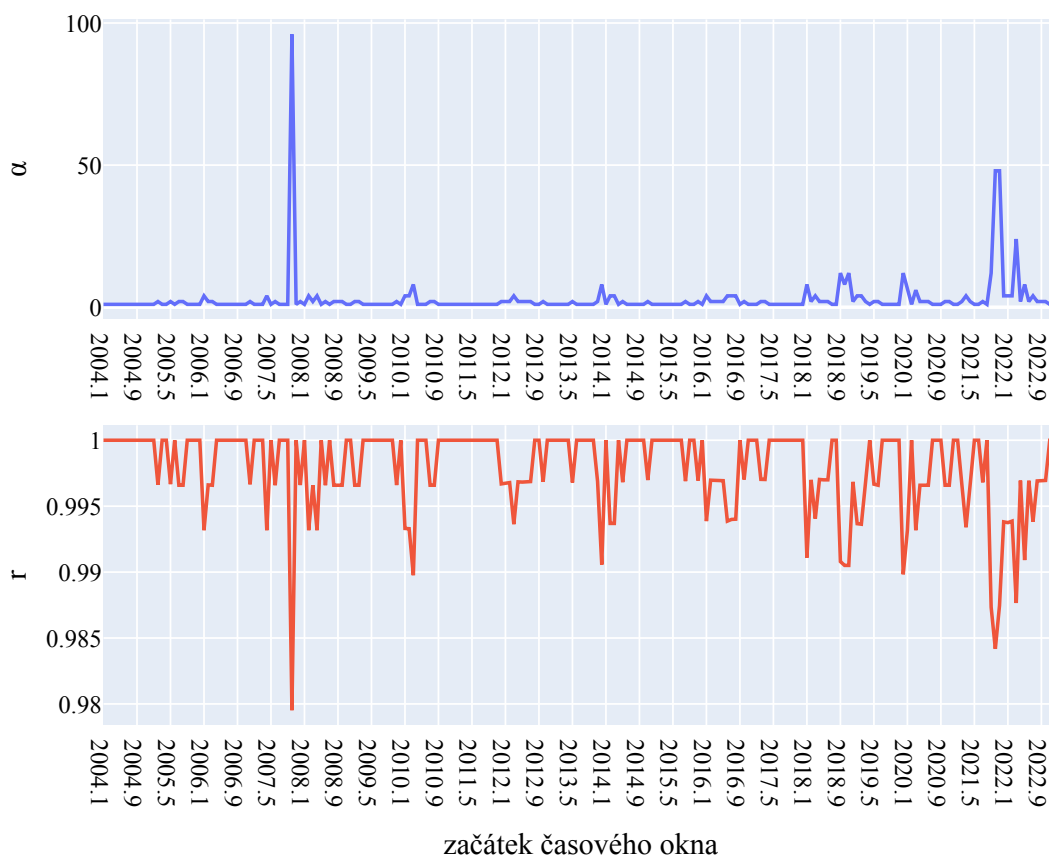
(AVA, NWN) – Utilities a 330 orbit velikosti 1. Síť má 338 vrcholů a její hustota je 0,212.

Dále vystupuje okno obsahující světovou finanční krizi (1. 3. 2008 až 1. 8. 2008), kde je velikost automorfnní grupy 48 a automorfismy tvoří 1 orbitu velikosti 4: (CLB, DRQ, HLX, SM) – Energy, 1 orbitu velikosti 2: (SNEX, TDS) a 312 orbit velikosti 1. Velikost sítě je 318 uzlů a její hustota je 0,24.

Co výstup této konfigurace kazí je další okno mající velikost automorfnní grupy 48 (1. 5. 2019 až 1. 10. 2019). Rozdíl je ve struktuře automorfismů, ty zde tvoří 1 orbitu velikosti 3: (AWR, CWT, NWN) – Utilities, 3 orbity velikosti 2: (CBRL, OMI), (CPK, MSEX) – Utilites, (MGEE, SJW) – Utilities a 323 orbit velikosti 1. Síť obsahuje 332 uzlů a hustotu má 0,22.

3.4 Shrnutí

Výše uvedené příklady ukazují, že výskyt symetrií ve Stock Network zřejmě není zcela náhodný. Niže uvádíme pár pozorování o výskytu konkrétních orbit. Z celkem 58 prozkoumaných orbit z nejlepších příkladů (orbity, které se objevily v obarveném i neobarveném grafu počítáme jen jednou) patřilo 26 do světové finanční krize, 16 do pandemie nemoci Covid-19 a 16 do prvního roku války na Ukrajině. Z těchto 58 orbit obsahovalo pouze 7 firmy z různých odvětví a z těchto 7 jich



Obrázek 3.7 Velikost automorfnní grupy α a normovaný počet orbit r největší komponenty souvislosti grafu sítě S&P 600 ohodnocené korelačním koeficientem v časových oknech velikosti 12 měsíců vytvořené metodou 3.

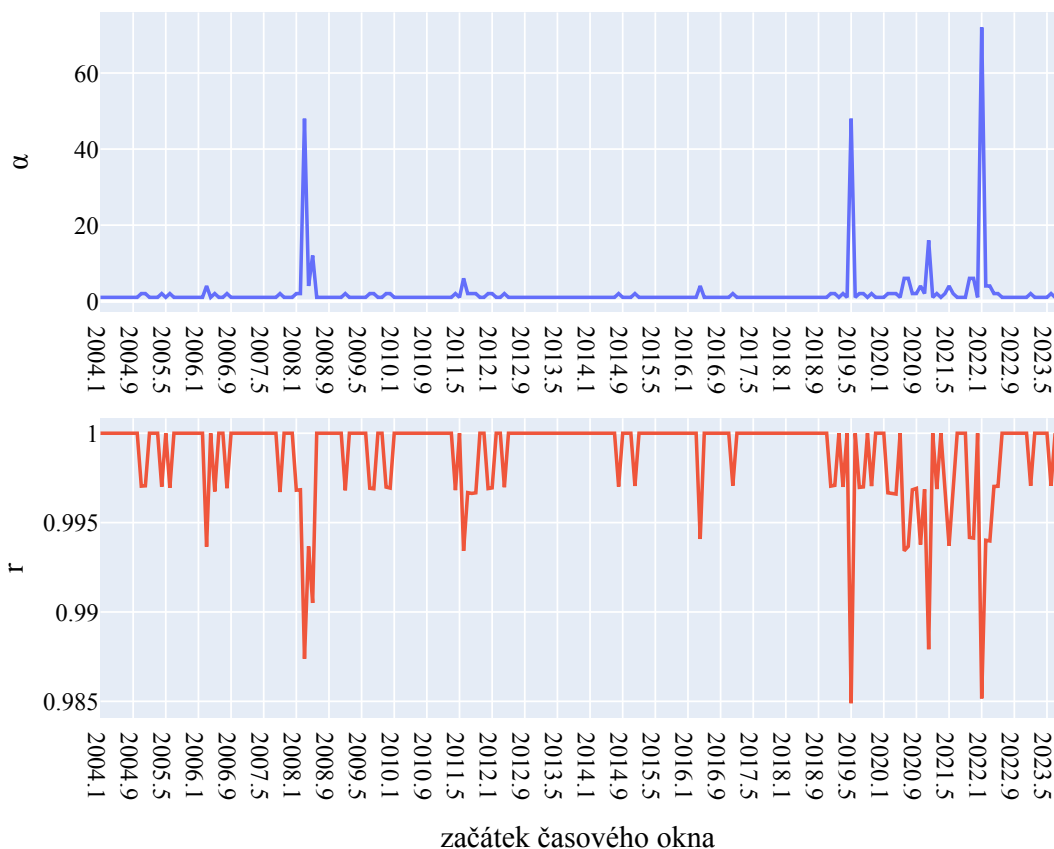
6 spadalo do světové finanční krize a 1 do pandemie nemoci Covid-19. Zajímavé je, že zatímco v krizích 2020 a 2022 se často objevovaly orbity tvořené firmami ze sektoru Utilities, v krizi 2008 taková orbita nebyla ani jedna. Na druhou stranu v krizích 2008 a 2022 se často objevovaly orbity tvořené firmami ze sektoru Energy, které se v krizi 2020 objevily pouze 2.

Jediná orbita, která se objevila třikrát, byla (CLB, DRQ, HLX, SM) – Energy a to vždy v časovém okně souvisejícím se světovou finanční krizí 2008. Konkrétně to bylo v obou příkladech metody 1: (1. 9. 2007 až 1. 8. 2008) v síti ohodnocené korelačním koeficientem a (1. 3. 2008 až 1. 8. 2008) v síti ohodnocené vzájemnou informací. Dále se objevila v tomtéž okně v síti vytvořené metodou 3 a ohodnocené vzájemnou informací.

Orbity, které se objevily dvakrát, byly 4. Jednalo se o (ETD, HVT) – Consumer Discretionary, která se pokaždé objevila v krizi 2020 a pokaždé v sítích vytvořených metodou 1. Konkrétně (1. 7. 2020 až 1. 6. 2021) v síti ohodnocené korelačním koeficientem a (1. 3. 2020 až 1. 8. 2020) v síti ohodnocené vzájemnou informací.

Další orbitou, která se objevila dvakrát byla (SVT.L, UU.L) – Utilities a to v oknech (1. 2. 2020 až 1. 4. 2020) a (1. 3. 2022 až 1. 5. 2022) v síti vytvořené metodou 2 a ohodnocené vzájemnou informací.

Třetí orbitou se dvěma výskyty byla (ATGE, STRA) – Consumer Discretionary. Ta se objevila pouze v sítích vytvořených metodou 1 a ohodnocených korelačním



Obrázek 3.8 Velikost automorfnní grupy α a normovaný počet orbit r největší komponenty souvislosti grafu sítě S&P 600 ohodnocené odhadem vzájemné informace s počtem binů 4 v časových oknech velikosti 6 měsíců vytvořené metodou 3.

koeficientem. Konkrétně v oknech (1. 9. 2007 až 1. 8. 2008) a (1. 10. 2007 až 1. 9. 2008).

Poslední takovou orbitou byla (SNEX, TDS), která je tvořena firmami z různých sektorů — Financials a Communication Services. Tato orbita se vyskytovala vždy v oknech spojených se světovou finanční krizí 2007 a 2008 a to v okně (1. 3. 2008 až 1. 8. 2008) v síti vytvořené metodou 1 ohodnocené vzájemnou informací a v tom samém okně v síti vytvořené metodou 3 opět ohodnocené vzájemnou informací.

Dále si lze všimnout, že některé akcie mají tendenci tvořit symetrie častěji než jiné. Nejvíce se ve sledovaných orbitách objevuje akcie se značkou CLB a to celkem v 7 — pětkrát v oknech týkajících se světové finanční krize 2007 a 2008: třikrát (CLB, DRQ, HLX, SM) – Energy, (CLB, HP, OII) – Energy, (CLB, CRK) – Energy a dvakrát v oknech týkajících se prvního roku války na Ukrajině: (CLB, RES) – Energy, (CLB, OII, RES) – Energy. Ačkoli se ve sledované množině navyskytuje orbita (CLB, OII) – Energy, dvojice firem CLB a OII se spolu vyskytuje ve dvou. To samé platí pro dvojici CLB a RES.

HLX se objevuje obdobně, ale v 6 orbitách — objevuje se pouze v jednom ze zkoumaných oken týkajících se války na Ukrajině. Během finanční krize 2007 a 2008 je to opět třikrát orbita (CLB, DRQ, HLX, SM) – Energy, dále (HLX, NBR, PTEN) – Energy a (HLX, HP, NBR) – Energy. V období prvního roku války na Ukrajině je to (DRQ, HLX, PTEN) – Energy. Můžeme si všimnout podobného

jevu, jako u prvku CLB, tedy že ačkoli některé dvojice netvoří vlastní orbitu, vyskytují se spolu v orbitách s dalším prvkem. Jedná se o dvojice HLX a NBR a HLX a PTEN.

Další prvky už se vyskytují v 5 a méně orbitách, nebudeme je uvádět. Lze si ale všimnout, že symetrie, orbity jimi tvořené a prvky v nich se nevyskytují náhodně, a že zřejmě opravdu reflektují události reálného světa.

4 Možnosti nahrazení symetrie

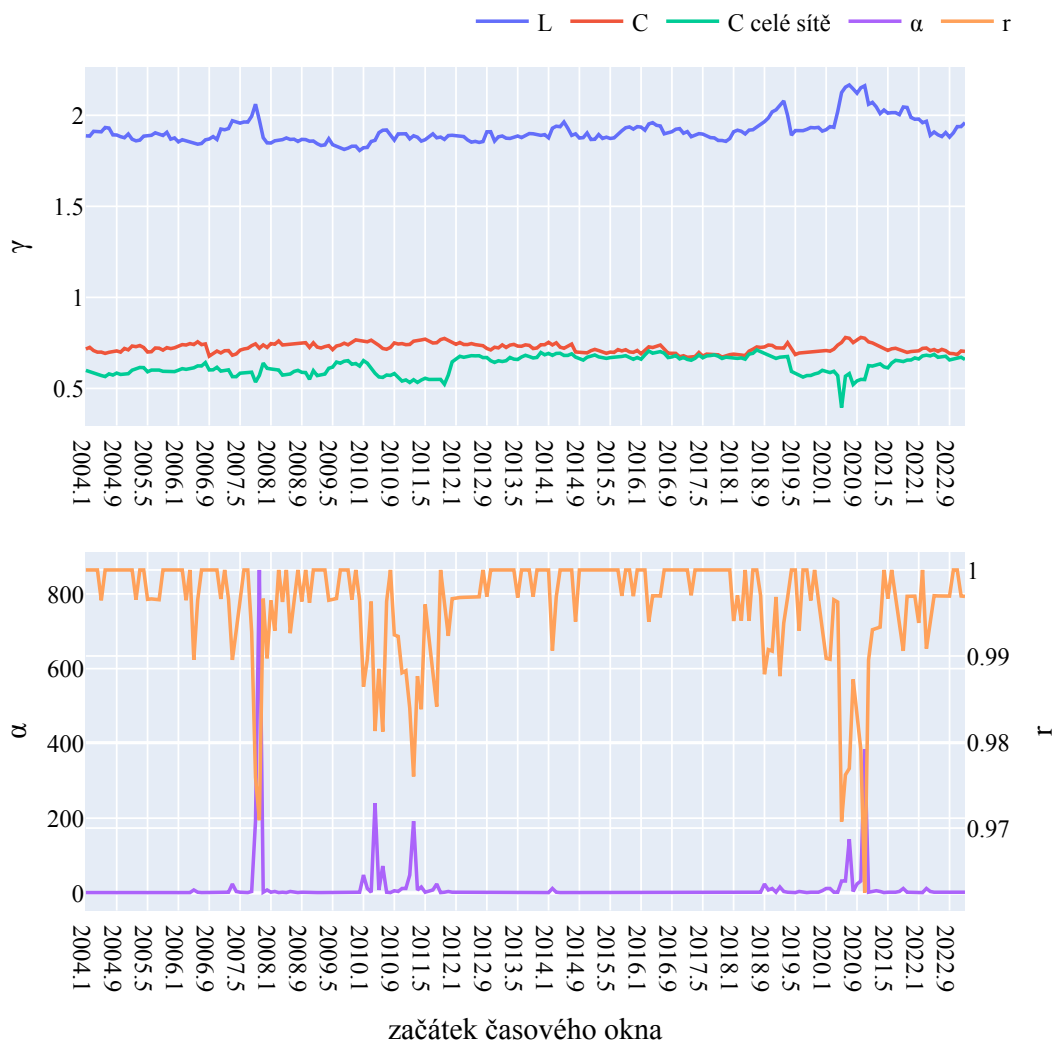
V této kapitole projdeme některé možnosti nahrazení symetrie jinou vlastností sítě.

4.1 Vlastnosti nesouvisející se symetrií

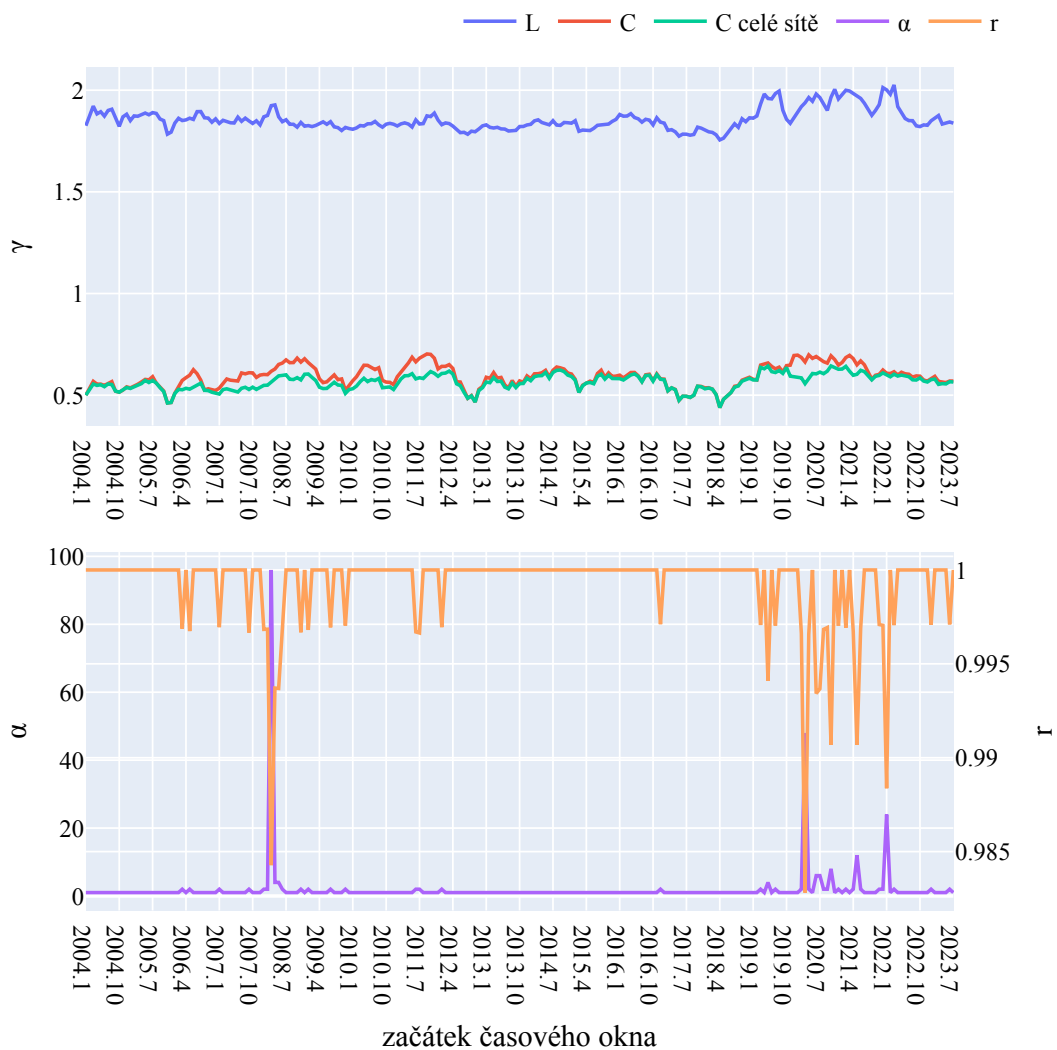
Čtenář by mohl namítnout, proč zkoumáme právě symetrie, které jsou — v porovnání s mnoha jinými vlastnostmi sítí — poměrně výpočetně náročné. V této sekci ukážeme, že symetrie nelze nahradit tradičními charakteristikami komplexních sítí, jako průměrnou nejkratší cestou a klastrovacím koeficientem. Další na seznamu by mohl být průměrný stupeň, ten je ale příliš závislý na hustotě (a naopak) a tedy jej neuvádíme. Pro naše nejlepší příklady ukážeme právě tyto veličiny, a to v případě klastrovacího koeficientu i na síti celé (na druhém grafu, který metody 1-3 vrací).

Na následujících stránkách si můžeme všimnout, že ani u jednoho z našich nejlepších příkladů nelze symetrii vyměnit za některý z uvedených ukazatelů. Dále si lze všimnout podobnosti mezi klastrovacím koeficientem největších komponent souvislosti a celých grafů. Větší odchylky jsou v časových oknech, kdy existovala nezanedbatelná část grafu, která byla mimo největší komponentu souvislosti — jinými slovy největší komponenta souvislosti byla relativně malá.

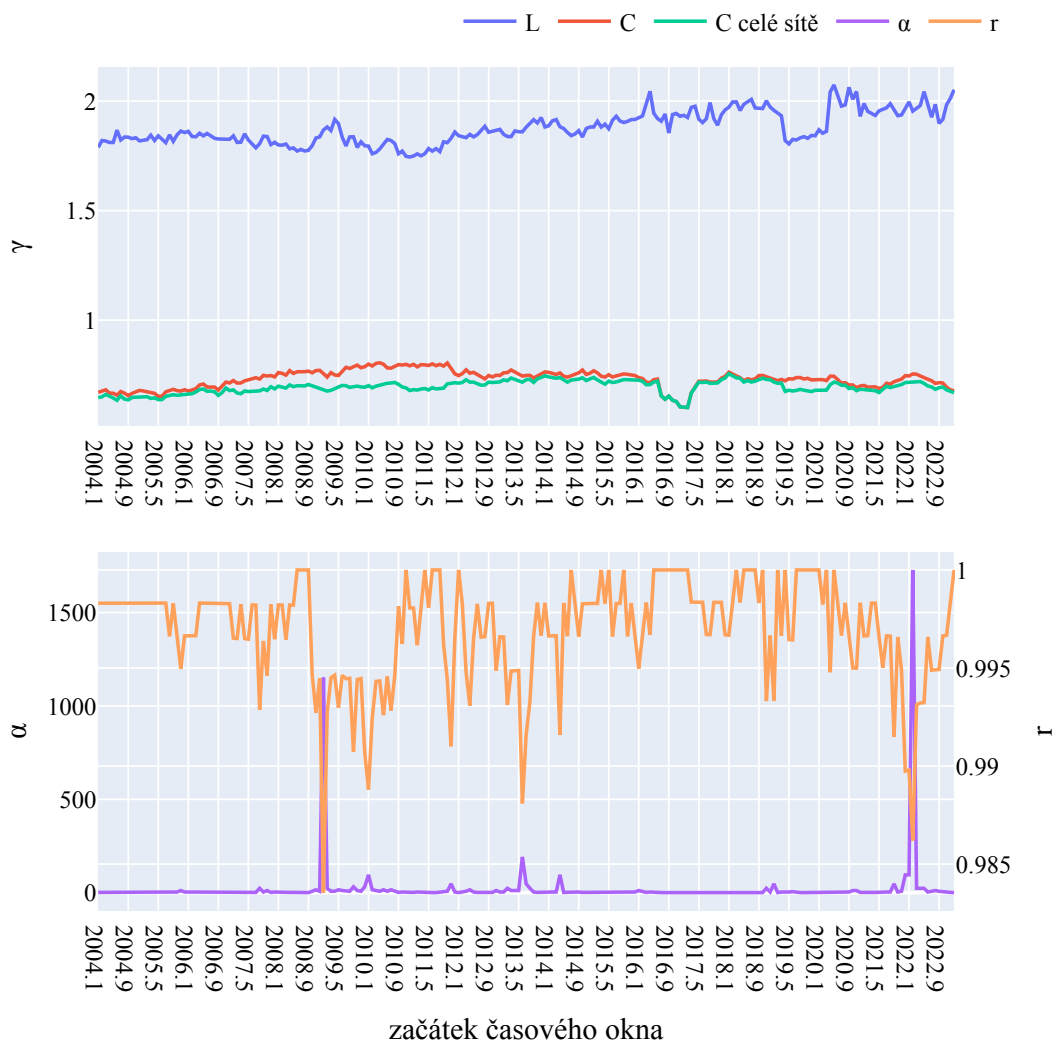
Závěrem této podkapitoly můžeme říct, že se nám symetrii průměrnou délkou nejkratší cesty ani klastrovacím koeficientem nahradit nepodařilo.



Obrázek 4.1 Vývoj průměrné délky nejkratší cesty L , klastrovacího koeficientu C a klastrovacího koeficientu C na celé síti v porovnání s velikostí automorfnní grupy α a normovaným počtem orbit r na síti tvořené nad indexem S&P 600 ohodnocené korelačním koeficientem a vytvořené metodou 1. Délka časového okna je 12 měsíců. Pro detailnější pohled na veličiny symetrie viz obrázek 3.1. Pro porovnání se symetrií respektující obarvení podle GICS viz obrázek 3.2.



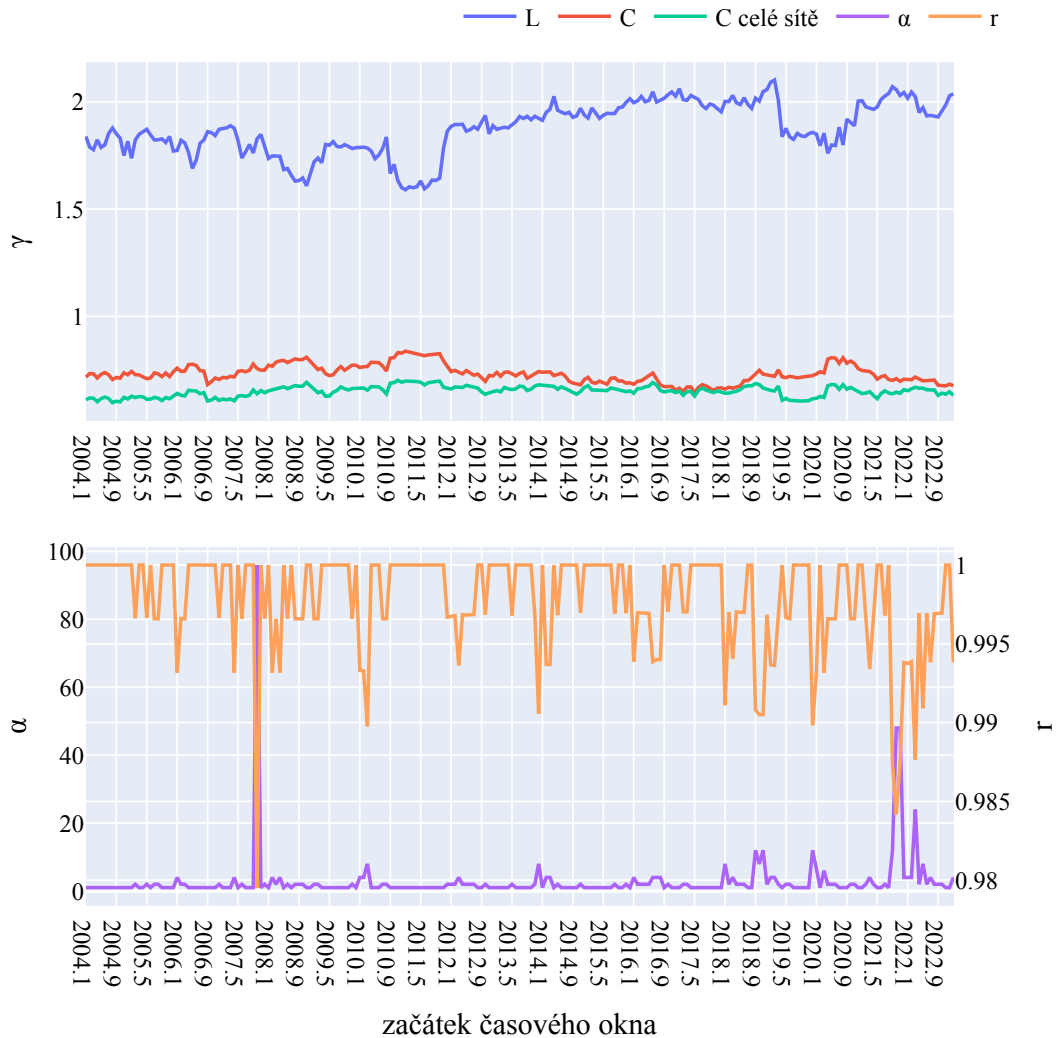
Obrázek 4.2 Vývoj průměrné délky nejkratší cesty L , klastrovacího koeficientu C a klastrovacího koeficientu C na celé síti v porovnání s velikostí automorfnní grupy α a normovaným počtem orbit r na síti tvořené nad indexem S&P 600 ohodnocené vzájemnou informací a vytvořené metodou 1. Délka časového okna je 6 měsíců a algoritmus pro odhad vzájemné informace je nastaven na počet binů 4. Pro detailnější pohled na veličiny symetrie viz obrázek 3.3.



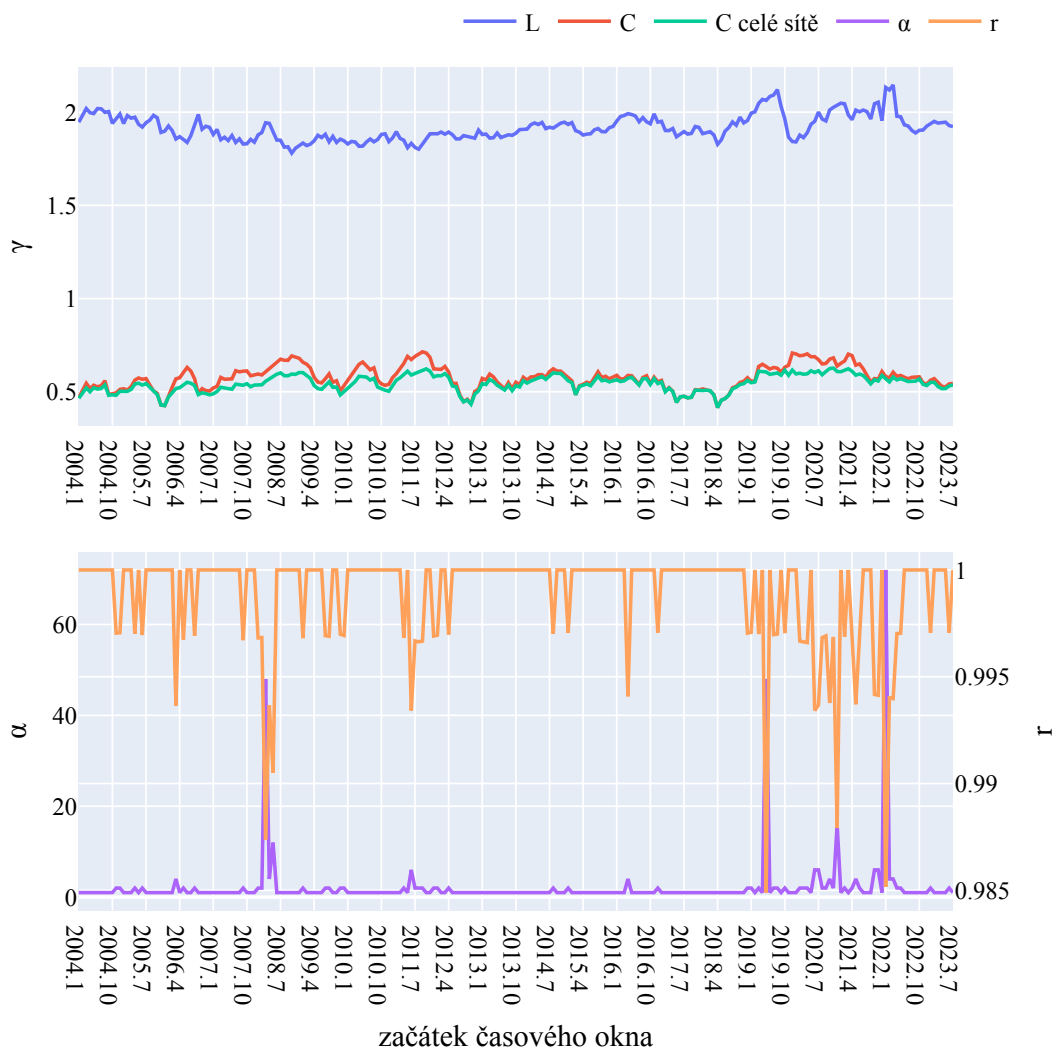
Obrázek 4.3 Vývoj průměrné délky nejkratší cesty L , klastrovacího koeficientu C a klastrovacího koeficientu C na celé síti v porovnání s velikostí automorfnní grupy α a normovaným počtem orbit r na síti tvořené nad indexem RUSSELL 1000 ohodnocené korelačním koeficientem a vytvořené metodou 2. Délka časového okna je 12 měsíců. Pro detailnější pohled na veličiny symetrie viz obrázek 3.4. Pro porovnání se symetrií respektující obarvení podle GICS viz obrázek 3.5.



Obrázek 4.4 Vývoj průměrné délky nejkratší cesty L , klastrovacího koeficientu C a klastrovacího koeficientu C na celé síti v porovnání s velikostí automorfnní grupy α a normovaným počtem orbit r na síti tvořené nad indexem FTSE 100 ohodnocené vzájemnou informací a vytvořené metodou 2. Délka časového okna jsou 3 měsíce, automorfismy respektují obarvení podle GICS a algoritmus pro odhad vzájemné informace je nastaven na počet binů 6. Pro detailnější pohled na veličiny symetrie viz obrázek 3.6.



Obrázek 4.5 Vývoj průměrné délky nejkratší cesty L , klastrovacího koeficientu C a klastrovacího koeficientu C na celé síti v porovnání s velikostí automorfnní grupy α a normovaným počtem orbit r na síti tvořené nad indexem S&P 600 ohodnocené korelačním koeficientem a vytvořené metodou 3. Délka časového okna je 12 měsíců. Pro detailnější pohled na veličiny symetrie viz obrázek 3.7.



Obrázek 4.6 Vývoj průměrné délky nejkratší cesty L , klastrovacího koeficientu C a klastrovacího koeficientu C na celé síti v porovnání s velikostí automorfní grupy α a normovaným počtem orbit r na síti tvořené nad indexem S&P 600 ohodnocené vzájemnou informací a vytvořené metodou 3. Délka časového okna je 6 měsíců a algoritmus pro odhad vzájemné informace je nastaven na počet binů 4. Pro detailnější pohled na veličiny symetrie viz obrázek 3.7.

4.2 Weisfeiler-Lehmanův test

Weisfeiler-Lehmanův test je iterační algoritmus, který dává jednosměrnou odpověď na otázku, zda jsou 2 grafy izomorfní. Pokud dá negativní odpověď, pak grafy určitě izomorfní nejsou. Pokud dá odpověď pozitivní, existuje dost velká šance, že grafy izomorfní jsou [29]. Lze jej tedy použít k odhadu symetričnosti grafu.

Algoritmus udržuje každému vrcholu hash, ten je určen podle sousedů daného vrcholu. Každá iterace znamená zohlednění další vrstvy sousedů vrcholu (tedy v první iteraci budou mít stejný hash vrcholy s právě stejným počtem sousedů každého stupně) a aktualizuje se v ní hash všech vrcholů.

K provedení WF-testu jsme použili funkci `weisfeiler_lehman_subgraph_hashes` z knihovny `NetworkX` [30]. Ta přiřadí každému z vrcholů hash a vrátí slovník, kde klíče jsou vrcholy a hodnoty jsou pole s hashem v každé iteraci. Vrcholy se stejným hashem ve stejné iteraci pak přiřadíme do stejné orbity. Takto získáme množinu všech aproximací orbit O a z počtu jejích prvků už pak přímo dostaneme normovaný počet orbit. Zároveň se znalostí množiny O můžeme odhadnout velikost automorfní grupy a to pomocí následujícího vzorce:

$$\alpha = \sum_{o \in O} |o|!$$

Kde $|o|$ označuje počet vrcholů v aproximaci orbity o .

V této sekci pustíme WF-test na našich nejlepších příkladech z kapitoly 3 a uvedeme výsledky po jedné a dvou iteracích. Celkově jsme WF-test spouštěli na 7 iteracích, viditelné rozdíly se ale objevily pouze mezi první a druhou. Lze si všimnout, že WF-test opravdu — oproti standardní symetrii — v některých případech automorfismy přidává, tedy co se týče první iterace. V druhé iteraci již vývoj velikosti automorfní grupy podle WF-testu téměř vždy odpovídá standardní symetrii. Jediné rozdíly se objevily u 2 našich nejlepších příkladů a to konkrétně u sítě ohodnocené korelačním koeficientem s největší komponentou získanou metodou 3:

- 1. 4. 2014 až 1. 3. 2014, velikost standardně získané automorfní grupy byla 1 a u grupy získané WF-testem byla velikost 2. Standardně získané orbity jsou pouze triviální — tedy velikosti 1 — zatímco WF-test našel jednu orbitu velikosti 2.
- 1. 9. 2021 až 1. 8. 2022, velikost standardně získané automorfní grupy byla 12 a u grupy získané WF-testem byla velikost 24. U obou způsobů dostáváme 2 orbity velikosti 2 a 1 orbitu velikosti 3.

A u sítě ohodnocené korelačním koeficientem a největší komponentou získanou metodou 1:

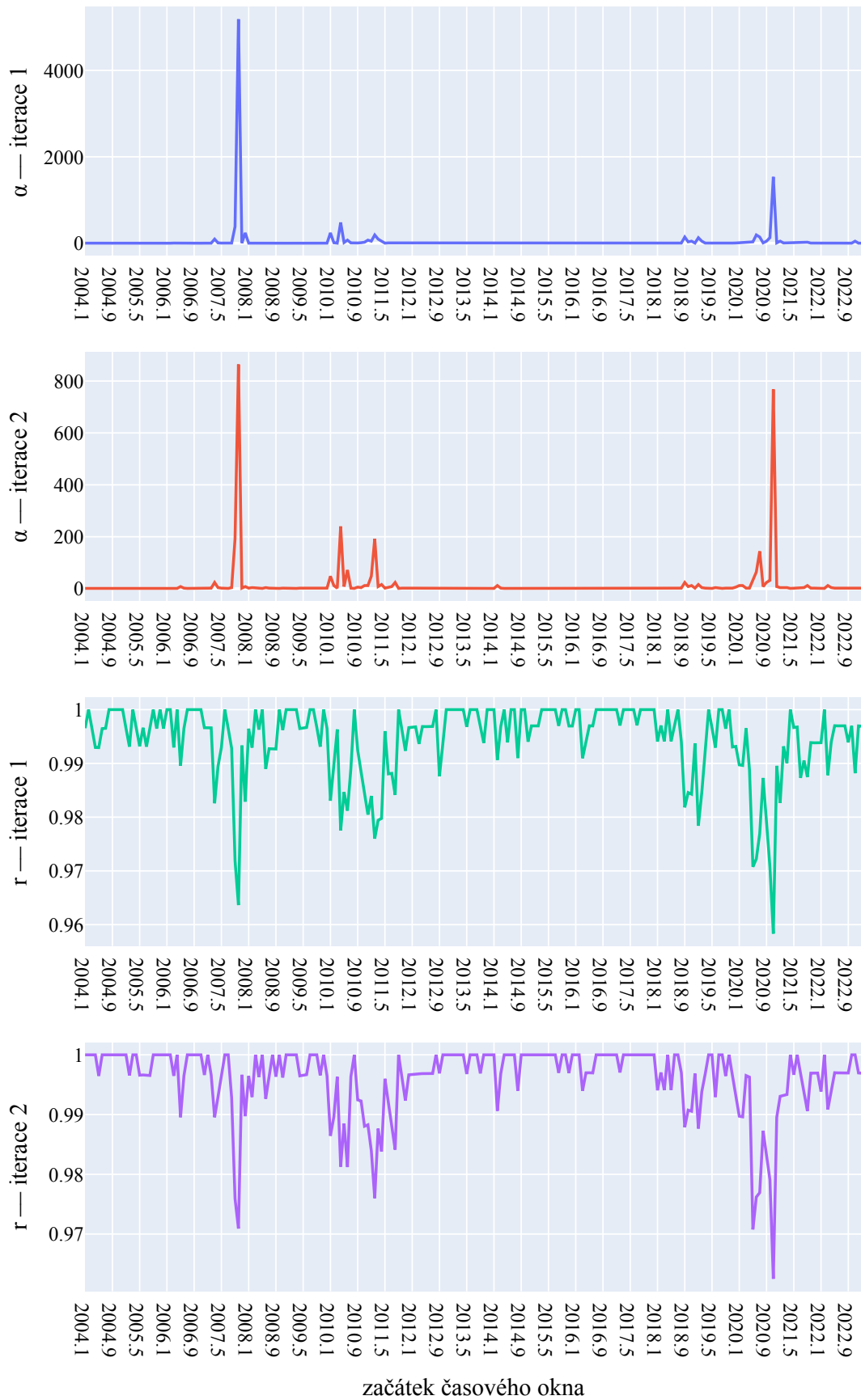
- 1. 6. 2020 až 1. 5. 2021, velikost standardně získané automorfní grupy byla 32 a u grupy získané WF-testem byla velikost 64. U obou způsobů dostáváme 5 orbit velikosti 2.
- 1. 8. 2020 až 1. 7. 2021, velikost standardně získané automorfní grupy byla 4 a u grupy získané WF-testem byla velikost 8. U obou způsobů dostáváme 3 orbity velikosti 2.

- 1. 11. 2020 až 1. 10. 2021, velikost standardně získané automorfní grupy byla 384 a u grupy získané WF-testem byla velikost 768. U obou způsobů dostáváme 7 orbit velikosti 2 a 1 orbitu velikosti 3.
- 1. 12. 2020 až 1. 11. 2021, velikost standardně získané automorfní grupy byla 2 a u grupy získané WF-testem byla velikost 8. U obou způsobů dostáváme 3 orbity velikosti 2.

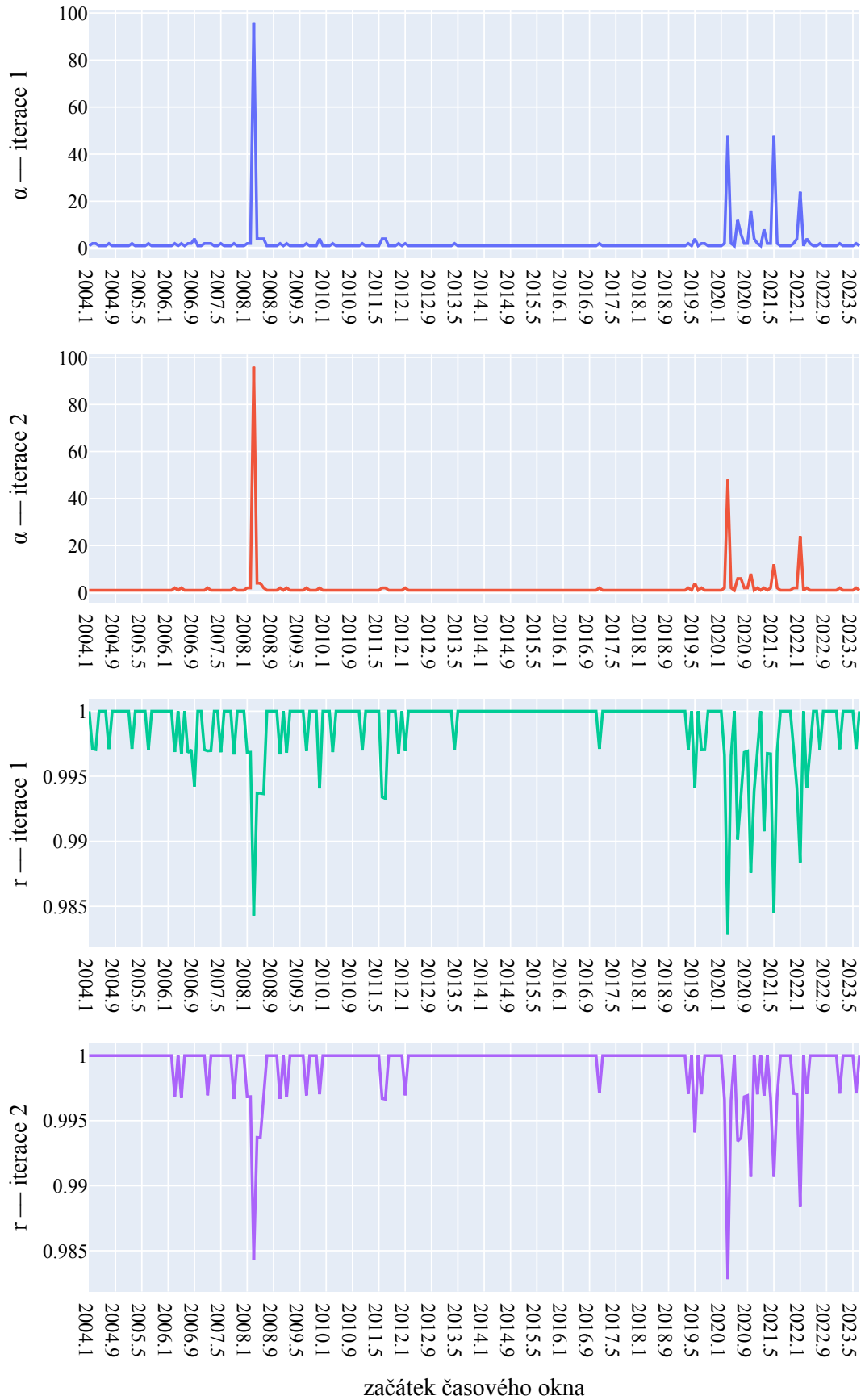
Všimněme si, že až na jeden případ — okno začínající 1. 12. 2020 je velikost automorfní grupy zjištěná WF-testem dvojnásobná oproti velikosti automorfní grupy získané standardním způsobem.

Až na jeden případ — okno začínající 1. 4. 2014 — nebylo špatné vyhodnocení velikosti automorfní grupy způsobeno nalezením falešné orbity.

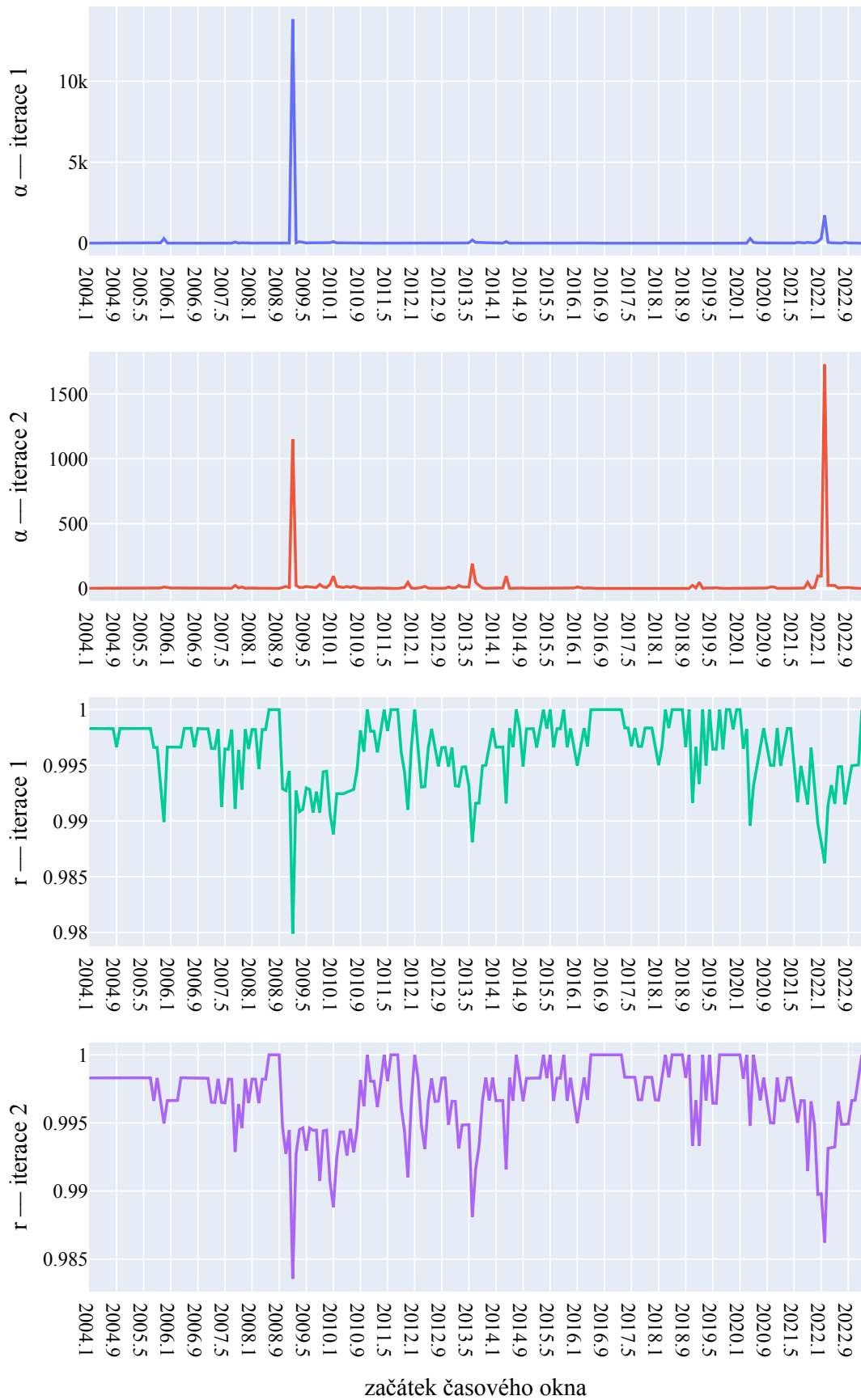
Závěrem této podkapitoly tedy můžeme říct, že symetrie na Stock Network lze odhadovat WF-testem po druhé iteraci. Ačkoli se WF-test od výsledků standardní symetrie ve dvou příkladech lišil, nezapříčinily tyto rozdíly zhoršení výsledku ve smyslu že by méně odpovídal událostem reálného světa.



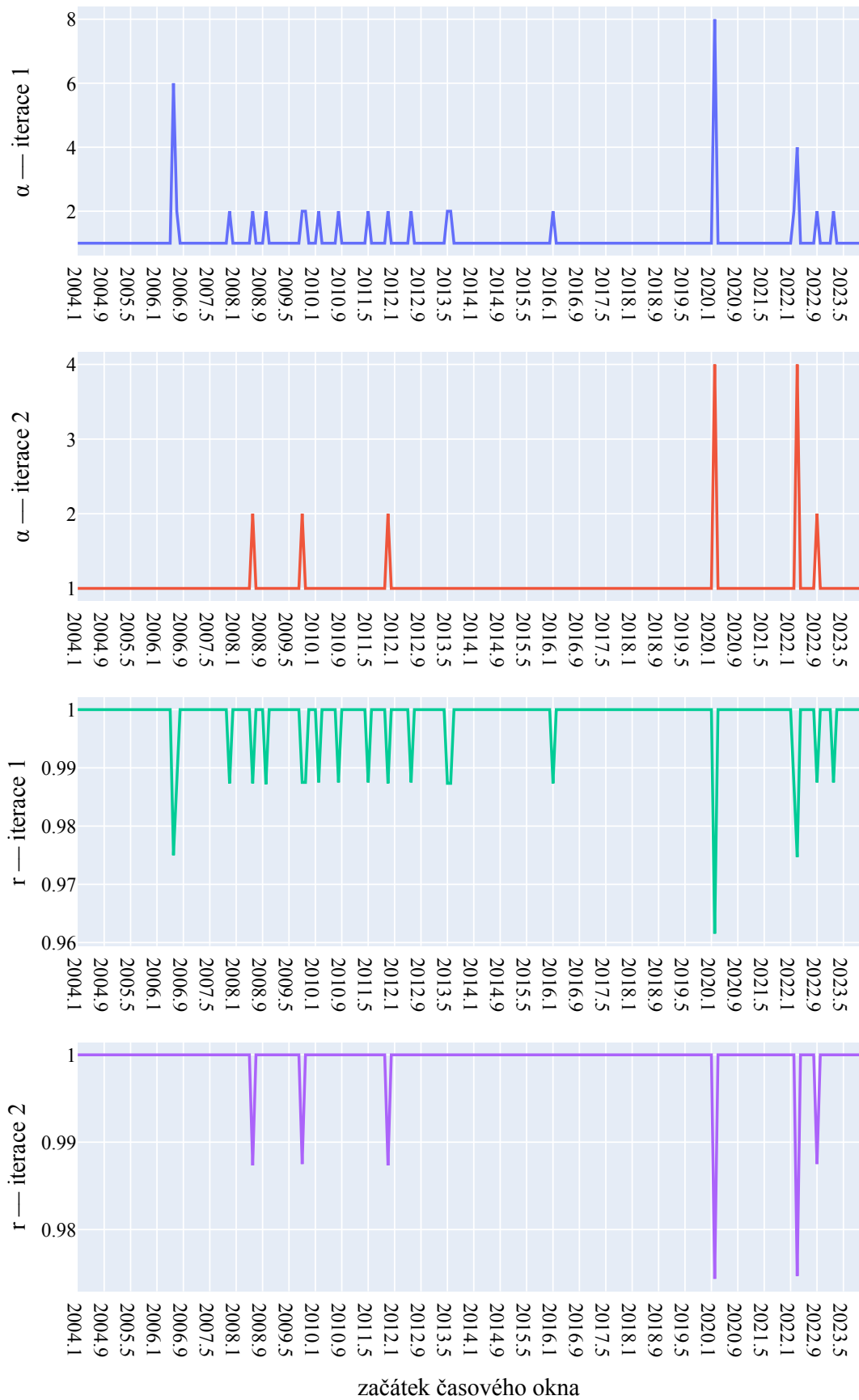
Obrázek 4.7 Odhad velikosti automorfní grupy normovaného počtu orbit pomocí WF-testu po první a druhé iteraci u příkladu sítě ohodnocené korelačním koeficientem u metody 1. Pro srovnání s hodnotami standardní symetrie viz obrázek 3.1.



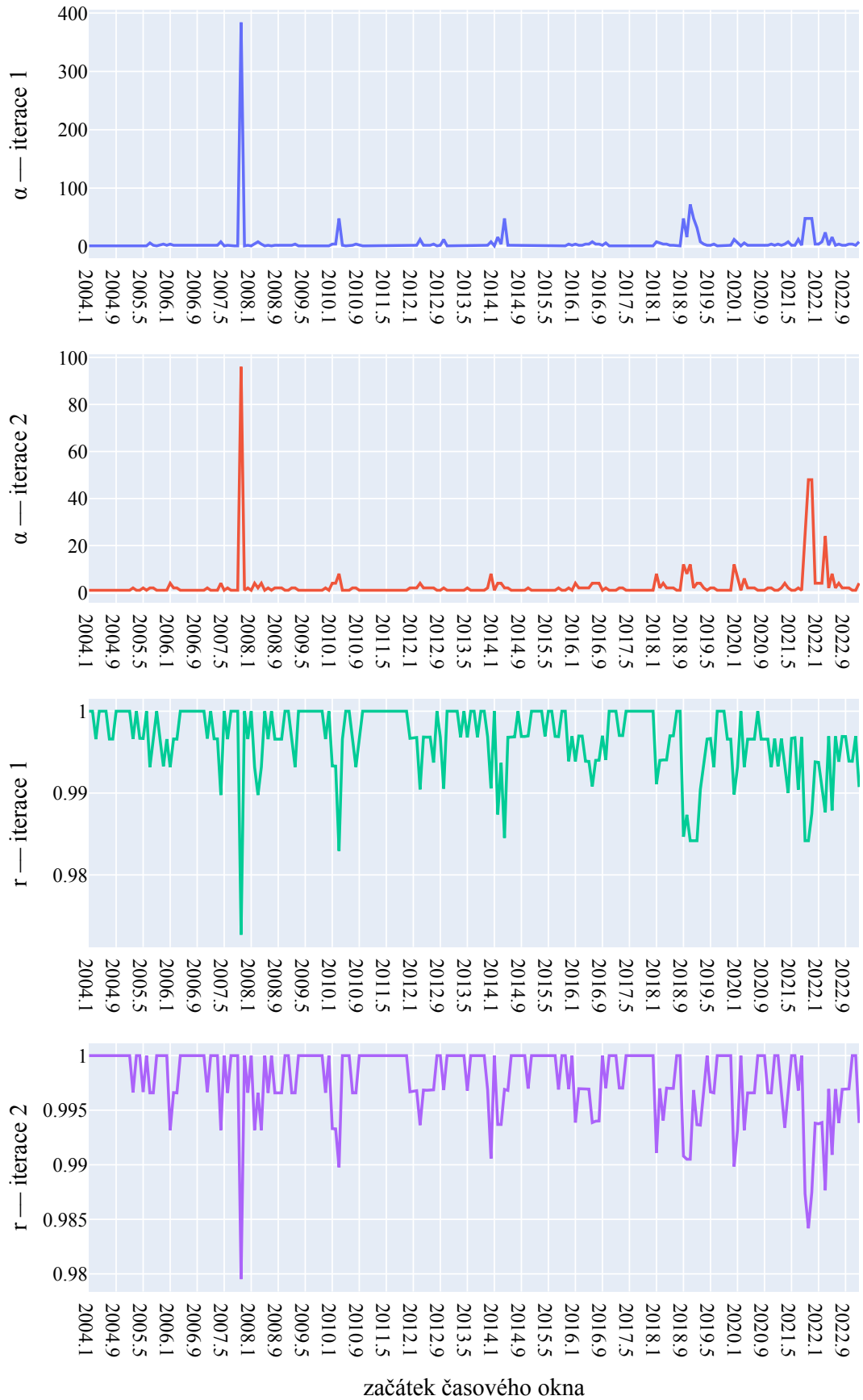
Obrázek 4.8 Odhad velikosti automorfní grupy normovaného počtu orbit pomocí WF-testu po první a druhé iteraci u příkladu sítě ohodnocené vzájemnou informací u metody 1. Pro srovnání s hodnotami standardní symetrie viz obrázek 3.3.



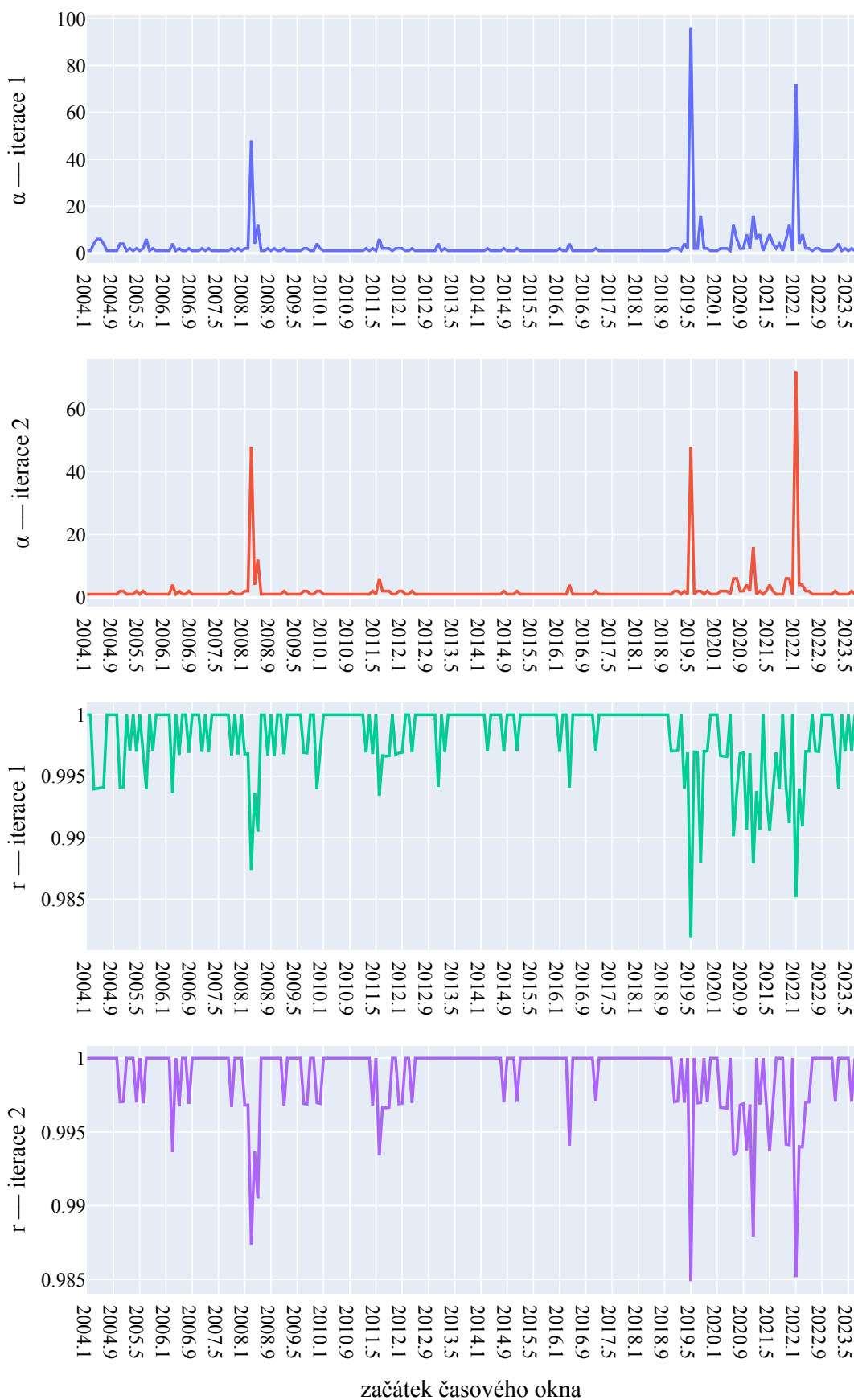
Obrázek 4.9 Odhad velikosti automorfní grupy normovaného počtu orbit pomocí WF-testu po první a druhé iteraci u příkladu sítě ohodnocené korelačním koeficientem u metody 2. Pro srovnání s hodnotami standardní symetrie viz obrázek 3.4.



Obrázek 4.10 Odhad velikosti automorfnní grupy normovaného počtu orbit pomocí WF-testu po první a druhé iteraci u příkladu sítě ohodnocené vzájemnou informací u metody 2. Druhá iterace vychází přesně, jako standardní symetrie bez respektování obarvení podle GICS. Pro srovnání s hodnotami standardní symetrie viz obrázek 3.6.



Obrázek 4.11 Odhad velikosti automorfnní grupy normovaného počtu orbit pomocí WF-testu po první a druhé iteraci u příkladu sítě ohodnocené korelačním koeficientem u metody 3. Pro srovnání s hodnotami standardní symetrie viz obrázek 3.7.



Obrázek 4.12 Odhad velikosti automorfnní grupy normovaného počtu orbit pomocí WF-testu po první a druhé iteraci u příkladu sítě ohodnocené vzájemnou informací u metody 3. Pro srovnání s hodnotami standardní symetrie viz obrázek 3.8.

Závěr

Předmětem této práce bylo zkoumání symetrií na Stock Network a jejich napojení na jevy reálného světa — světovou finanční krizi let 2007 a 2008, vypuknutí pandemie nemoci Covid-19 a následné uzavření celosvětové ekonomiky a první rok války na Ukrajině. Stock Network jsme stavěli nad 5 burzovními idnexy, z nichž 3 jsou definovány na americkém a 2 na evropském trhu. Stock Network jsme ohodnotili jak korelačním koeficientem, tak vzájemnou informací, kterou jsme odhadovali algoritmem EQQ, který jsme spouštěli s různými hodnotami parametru počtu binů.

Představili jsme 3 metody pro získání grafu sítě k analýze symetrie, které na rozdíl od běžně používaných metod — maximální kostry grafu, maximálního rovinného filtrovaného grafu a strategie vítěz bere vše — nesnižují drasticky počet hran, zachovávají pořadí hran podle jejich ohodnocení a vracejí souvislý graf, resp. graf s rozumně velkou největší komponentou.

Na grafech takto získaných sítí — celkově 165 — jsme spustili analýzu symetrie to jak bez, tak s ohledem na obarvení podle industriálních sektorů GICS. Symetrii jsme měřili velikostí automorfní grupy a normovaným počtem orbit. Prezentovali jsme 6 nejlepších výsledků, u kterých jsme popsali velikost automorfní grupy a distribuci počtu orbit v časových oknech odpovídajících sledovaným událostem reálného světa.

Také jsme se podívali na složení těchto orbit a zjistili jsme, že nejčastěji v nich figurují firmy působící v sektoru Energy, Utilities a Consumer Discretionary. V oknech náležících k finanční krizi 2007 a 2008 se symetrii nejčastěji účastnily firmy se značkami CLB a HLX (obě působící v sektoru Energy). V oknech týkajících se pandemie nemoci Covid-19 se v orbitách nejčastěji objevovala firma se značkou HVT působící v sektoru Consumer Discretionary. V oknech týkajících se prvního roku války na Ukrajině v orbitách nejčastěji byly firmy se značkami AVA a RES. První působí v sektoru Utilities, druhá v sektoru Energy.

U nejlepších příkladů jsme dále vyzkoušeli symetrii nahradit průměrnou nejkratší cestou a klastrovacím koeficientem a ukázali jsme, že takto symetrii nahradit nelze.

Také jsme symetrii vyzkoušeli nahradit Weisfeiler-Lehmanovým testem, který jsme spustili na 7 iterací a zjistili jsme, že v našem způsobem získaných sítích jsou rozdíly pouze mezi iterací první a druhou. Pomocí WF-testu jsme odhadli velikost automorfní grupy. Výsledky druhé iterace odhadly velikost automorfní grupy na téměř všech našich nejlepších příkladech správně. Jediné rozdíly se objevily u sítě vytvořené metodou 1 a ohodnocené korelačním koeficientem a u sítě vytvořené metodou 3, taktéž ohodnocené korelačním koeficientem. Obě tyto sítě měly velikost časového okna nastavenou na 12 měsíců a obě jsou tvořeny nad idnexem S&P 600. Rozdíly jsme detailně popsali v podkapitole 4.2. U většiny se nejednalo o nalezení falešné orbity.

4.3 Prostor pro další práci

V práci jsme na několika příkladech ukázali, že se události reálného světa propisují na symetrii vhodně postavené Stock Network. Dalším krokem by mohlo

být bližší specifikování jak správně Stock Network postavit, aby události reálného světa refletovala co nejlépe. Ukázali jsme, že symetrie nelze nahradit průměrnou nejkratší cestou a klastrovacím koeficientem, dalším krokem by mohlo být prozkoumání dalších charakteristik. Symetrii jsme zkoušeli odhadnout WF-testem, jehož výsledky na 2 z 8 nejlepších příkladů obsahovaly nepřesnosti. Dalším krokem by mohlo být nalezení příčiny těchto nepřesností.

Literatura

1. WATTS, D. J.; STROGATZ, S. H. Collective dynamics of ‘small-world’ networks. *nature*. 1998, roč. 393, č. 6684, s. 440.
2. SANCHEZ-GARCIA, D. R. Exploiting symmetry in network analysis. *Communications Physics*. 2020, roč. 3, č. 1, s. 2–3.
3. PAPO, D.; BULDÚ, J. M.; BOCCALETI, S.; BULLMORE, E. T. Complex network theory and the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2014, roč. 369, č. 1653, s. 20130520.
4. WILLIAMS, R.; MARTINEZ, N. Simple Rules Yield Complex Food Webs. *Nature*. 2000, roč. 404, č. 6774, s. 1800–183.
5. SIGANOS, G.; FALOUTSOS, M.; FALOUTSOS, P.; FALOUTSOS, C. Power laws and the AS-level Internet topology. *IEEE/ACM Transactions on Networking*. 2003, roč. 11, č. 4, s. 514–524.
6. WEYL, H. Symmetry. In: Princeton University Press, 1952, kap. Preface.
7. MACARTHUR, B. D.; SÁNCHEZ-GARCÍA, R. J.; ANDERSON, J. W. Symmetry in complex networks. *Discrete Applied Mathematics*. 2008, roč. 156, č. 18, s. 3525–3531.
8. WANG, H.; YAN, G.; XIAO, Y. Symmetry in world trade network. *Journal of Systems Science and Complexity*. 2009, roč. 22, č. 2, s. 280–290.
9. MANTEGNA, R. N. Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*. 1999, roč. 11, s. 193–197.
10. CHI, K. T.; LIU, J.; LAU, F. C. M. A network perspective of the stock market. *Journal of Empirical Finance*. 2010, roč. 17, č. 4, s. 659–667.
11. TUMMINELLO, M.; ASTE, T.; DI MATTEO, T.; MANTEGNA, R. N. A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences*. 2005, roč. 102, č. 30, s. 10421–10426.
12. MATOUŠEK, J.; NEŠETŘIL, J. *Kapitoly z diskrétní matematiky*. Praha: Karolinum, 2002. Druhé opravené vydání. ISBN 80-246-0084-6.
13. BALL, F.; GEYER-SCHULZ, A. How Symmetric Are Real-World Graphs? A Large-Scale Study. *Symmetry*. 2018, roč. 10, č. 1, s. 2–7.
14. KLIN, M.; RÜCKER, Ch.; RÜCKER, G.; TINHOFER, G. Algebraic Combinatorics in Mathematical Chemistry, Methods and Algorithms, I. Permutation Groups and Coherent (Cellular) Algebras. In: Technische Universität München, 1995, kap. 4.
15. BARABÁSI, A. L.; BONABEAU, E. Scale-Free Networks. *Scientific American*. 2003, roč. 288, č. 5, s. 60–62.
16. WANG, X. F.; CHEN, G. Complex networks: Small-world, scale-free and beyond. *IEEE Circuits and Systems Magazine*. 2003, roč. 3, č. 1, s. 6–20.
17. WIKIMEDIA FOUNDATION. *List of S&P 500 companies*. 2024-04-01. Dostupné také z: https://en.wikipedia.org/wiki/List_of_S%26P_500_companies.

18. WIKIMEDIA FOUNDATION. *List of S&P 600 companies*. 2024-04-01. Dostupné také z: https://en.wikipedia.org/wiki/List_of_S%26P_600_companies.
19. WIKIMEDIA FOUNDATION. *FTSE 100 Index*. 2024-04-01. Dostupné také z: https://en.wikipedia.org/wiki/FTSE_100_Index.
20. WIKIMEDIA FOUNDATION. *FTSE 250 Index*. 2024-04-01. Dostupné také z: https://en.wikipedia.org/wiki/FTSE_250_Index.
21. WIKIMEDIA FOUNDATION. *Russell 1000 Index*. 2024-04-01. Dostupné také z: https://en.wikipedia.org/wiki/Russell_1000_Index.
22. YAHOO FINANCE. *Yahoo Finance*. 2024-04-01. Dostupné také z: <https://finance.yahoo.com>.
23. JAJCAY, N. *pyCliTS*. 2024-04-01. Dostupné také z: <https://github.com/jajcayn/pyclits/tree/master>.
24. HARTMAN, D.; HLINKA, J. Nonlinearity in stock networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*. 2018, roč. 28, č. 8, s. 083127.
25. MSCI. *The Global Industry Classification Standard (GICS®)*. 2024-05-01. Dostupné také z: <https://www.msci.com/our-solutions/indexes/gics>.
26. M., Škola F. *Stock-network analysis*. 2024-04-01. Dostupné také z: <https://gitlab.mff.cuni.cz/teaching/nprg045/hart%20man/frantisek-skola.git>.
27. THE SAGE DEVELOPERS. *SageMath, the Sage Mathematics Software System (Version 10.1.)* 2024-04-01. Dostupné také z: <https://www.sagemath.org/index.html>.
28. THE PYDATA DEVELOPMENT TEAM. *pandas-datareader*. 2024-04-01. Dostupné také z: <https://pandas-datareader.readthedocs.io/en/latest/>.
29. HUANG, Ningyuan Teresa; VILLAR, Soledad. A Short Tutorial on The Weisfeiler-Lehman Test And Its Variants. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2021.
30. HAGBERG, Aric A.; SCHULT, Daniel A.; SWART, Pieter J. Exploring Network Structure, Dynamics, and Function using NetworkX. In: VAROQUAUX, Gaël; VAUGHT, Travis; MILLMAN, Jarrod (ed.). *Proceedings of the 7th Python in Science Conference*. Pasadena, CA USA, 2008, s. 11–15.

Seznam obrázků

2.1	Počet vrcholů největší komponenty souvislosti grafu sítě FTSE 100 ohodnocené korelací v časových oknech velikosti 3 měsíce vytvořené metodou 1.	17
2.2	Počet vrcholů největší komponenty souvislosti grafu sítě FTSE 100 ohodnocené korelací v časových oknech velikosti 3 měsíce vytvořené metodou 2.	17
2.3	Počet vrcholů největší komponenty souvislosti grafu sítě FTSE 100 ohodnocené korelací v časových oknech velikosti 3 měsíce vytvořené metodou 3.	17
3.1	Velikost automorfnní grupy α a normovaný počet orbit r největší komponenty souvislosti grafu sítě S&P 600 ohodnocené korelačním koeficientem v časových oknech velikosti 12 měsíců vytvořené metodou 1.	21
3.2	Velikost automorfnní grupy α a normovaný počet orbit r respektující obarvení podle GICS největší komponenty souvislosti grafu sítě S&P 600 ohodnocené korelačním koeficientem v časových oknech velikosti 12 měsíců vytvořené metodou 1.	22
3.3	Velikost automorfnní grupy α a normovaný počet orbit r největší komponenty souvislosti grafu sítě S&P 600 ohodnocené vzájemnou informací s počtem binů 4 v časových oknech velikosti 6 měsíců vytvořené metodou 1.	23
3.4	Velikost automorfnní grupy α a normovaný počet orbit r komponenty souvislosti grafu sítě RUSSELL 1000 ohodnocené korelačním koeficientem v časových oknech velikosti 12 měsíců vytvořené metodou 2.	24
3.5	Velikost automorfnní grupy α a normovaný počet orbit r respektující obarvení největší komponenty souvislosti grafu sítě RUSSELL 1000 ohodnocené korelačním koeficientem v časových oknech velikosti 12 měsíců vytvořené metodou 2.	25
3.6	Velikost automorfnní grupy α a normovaný počet orbit r respektující obarvení největší komponenty souvislosti grafu sítě FTSE 100 ohodnocené vzájemnou informací s počtem binů 6 v časových oknech velikosti 3 měsíce vytvořené metodou 2.	26
3.7	Velikost automorfnní grupy α a normovaný počet orbit r největší komponenty souvislosti grafu sítě S&P 600 ohodnocené korelačním koeficientem v časových oknech velikosti 12 měsíců vytvořené metodou 3.	27
3.8	Velikost automorfnní grupy α a normovaný počet orbit r největší komponenty souvislosti grafu sítě S&P 600 ohodnocené odhadem vzájemné informace s počtem binů 4 v časových oknech velikosti 6 měsíců vytvořené metodou 3.	28

4.1	Vývoj průměrné délky nejkratší cesty L , klastrovacího koeficientu C a klastrovacího koeficientu C na celé síti v porovnání s velikostí automorfnní grupy α a normovaným počtem orbit r na síti tvořené nad indexem S&P 600 ohodnocené korelačním koeficientem a vytvořené metodou 1. Délka časového okna je 12 měsíců. Pro detailnější pohled na veličiny symetrie viz obrázek 3.1. Pro porovnání se symetrií respektující obarvení podle GICS viz obrázek 3.2.	31
4.2	Vývoj průměrné délky nejkratší cesty L , klastrovacího koeficientu C a klastrovacího koeficientu C na celé síti v porovnání s velikostí automorfnní grupy α a normovaným počtem orbit r na síti tvořené nad indexem S&P 600 ohodnocené vzájemnou informací a vytvořené metodou 1. Délka časového okna je 6 měsíců a algoritmus pro odhad vzájemné informace je nastaven na počet binů 4. Pro detailnější pohled na veličiny symetrie viz obrázek 3.3.	32
4.3	Vývoj průměrné délky nejkratší cesty L , klastrovacího koeficientu C a klastrovacího koeficientu C na celé síti v porovnání s velikostí automorfnní grupy α a normovaným počtem orbit r na síti tvořené nad indexem RUSSELL 1000 ohodnocené korelačním koeficientem a vytvořené metodou 2. Délka časového okna je 12 měsíců. Pro detailnější pohled na veličiny symetrie viz obrázek 3.4. Pro porovnání se symetrií respektující obarvení podle GICS viz obrázek 3.5.	33
4.4	Vývoj průměrné délky nejkratší cesty L , klastrovacího koeficientu C a klastrovacího koeficientu C na celé síti v porovnání s velikostí automorfnní grupy α a normovaným počtem orbit r na síti tvořené nad indexem FTSE 100 ohodnocené vzájemnou informací a vytvořené metodou 2. Délka časového okna jsou 3 měsíce, automorfismy respektují obarvení podle GICS a algoritmus pro odhad vzájemné informace je nastaven na počet binů 6. Pro detailnější pohled na veličiny symetrie viz obrázek 3.6.	34
4.5	Vývoj průměrné délky nejkratší cesty L , klastrovacího koeficientu C a klastrovacího koeficientu C na celé síti v porovnání s velikostí automorfnní grupy α a normovaným počtem orbit r na síti tvořené nad indexem S&P 600 ohodnocené korelačním koeficientem a vytvořené metodou 3. Délka časového okna je 12 měsíců. Pro detailnější pohled na veličiny symetrie viz obrázek 3.7.	35
4.6	Vývoj průměrné délky nejkratší cesty L , klastrovacího koeficientu C a klastrovacího koeficientu C na celé síti v porovnání s velikostí automorfnní grupy α a normovaným počtem orbit r na síti tvořené nad indexem S&P 600 ohodnocené vzájemnou informací a vytvořené metodou 3. Délka časového okna je 6 měsíců a algoritmus pro odhad vzájemné informace je nastaven na počet binů 4. Pro detailnější pohled na veličiny symetrie viz obrázek 3.7.	36
4.7	Odhad velikosti automorfnní grupy normovaného počtu orbit pomocí WF-testu po první a druhé iteraci u příkladu sítě ohodnocené korelačním koeficientem u metody 1. Pro srovnání s hodnotami standardní symetrie viz obrázek 3.1.	39

4.8	Odhad velikosti automorfní grupy normovaného počtu orbit pomocí WF-testu po první a druhé iteraci u příkladu sítě ohodnocené vzájemnou informací u metody 1. Pro srovnání s hodnotami standardní symetrie viz obrázek 3.3.	40
4.9	Odhad velikosti automorfní grupy normovaného počtu orbit pomocí WF-testu po první a druhé iteraci u příkladu sítě ohodnocené korelačním koeficientem u metody 2. Pro srovnání s hodnotami standardní symetrie viz obrázek 3.4.	41
4.10	Odhad velikosti automorfní grupy normovaného počtu orbit pomocí WF-testu po první a druhé iteraci u příkladu sítě ohodnocené vzájemnou informací u metody 2. Druhá iterace vychází přesně, jako standardní symetrie bez respektování obarvení podle GICS. Pro srovnání s hodnotami standardní symetrie viz obrázek 3.6.	42
4.11	Odhad velikosti automorfní grupy normovaného počtu orbit pomocí WF-testu po první a druhé iteraci u příkladu sítě ohodnocené korelačním koeficientem u metody 3. Pro srovnání s hodnotami standardní symetrie viz obrázek 3.7.	43
4.12	Odhad velikosti automorfní grupy normovaného počtu orbit pomocí WF-testu po první a druhé iteraci u příkladu sítě ohodnocené vzájemnou informací u metody 3. Pro srovnání s hodnotami standardní symetrie viz obrázek 3.8.	44

Seznam tabulek

2.1	Skutečně použité počty akcií.	13
2.2	Tabulka všech parametrů.	14
2.3	Počty skutečně použitých akcií podle industriálních sektorů GICS. 1–Energy, 2–Materials, 3–Industrials, 4–Consumer Discretionary, 5–Consumer Staples, 6–Health Care, 7–Financials, 8–Information Tech- nology, 9–Communication Services, 10–Utilities, 11–Real Estate .	19