1. 4 facts
2. 3-axes → communication
3. **CSCW**
4. N-cliques → why N
5. clique-overlap
6. can you solve G(V, E) cliques efficiently?
7. SPB → delta_ab explain
8. Stress… → alpha_ab
9. **what's the difference between Game & Decision theory**
10. NE → under condition
11. what is AUs
12. different types of silence
13. GMM
14. greedy
15. fuzzy-c-mean
16. single/complete linkage
17. feedback centrality
18. differential attachment
19. **difference between facial detection and facial recognition**
20. **Authority and Hub**
21. **Homophily**

Social Games

- Name two scientific methodologies that ludologists may use! Name four scientific fields that make contributions to ludology!
  - Empirical 經驗 & Deductive 演繹
  - Economy & Informatics & Psychology & Arts

- Name four characteristics of “play” following Huizinga or Caillois!
  - Free
  - outside “ordinary” life, as being “not serious”
  - fixed rules
  - formation of social groupings

- Briefly explain the classes Agon, Alea, Mimicry and Illinx in Caillois’ taxonomy of play and games!
  - Agon → Competition
    - Not regulated → Racing, Wrestling, Etc.
    - In general → Contest, Sports
  - Alea → Chance
    - counting-out, rhymes, Heads or tails
    - simple, complex, continuing, lotteries
  - Mimicry → Simulation
    - Children’s initiations, Game of illusion, Tag, Arms, Masks
    - Theater, Spectacles in general
  - Ilinx → Vertigo
    - Children’s “whirling”, horseback riding, swinging

- Salen and Zimmerman state: “Games are a subset of play. Play is an element of games” Briefly discuss the meaning and difficulties of that definition!
  - play (Activity) → Formalized interaction when players follow the rules of a game and experience its system through play
  - game (Activity or Object)→ Is a **system**, in which **players engage** in

- Salen and Zimmerman define:
  “A game is a system in which players engage in an artificial conflict, defined by rules, that results in a quantifiable outcome.”
  Discuss the limits of this definition!
  - systems that **build behavior via interaction**
  - emergent property of the game as defined by its rules
  - what about the **digital games**?

- What do Salen and Zimmerman understand by “transformative social play”?
    - “In transformative social play players extend, transform, and manipulate existing social relationships through play itself”

- Characterization of player types: explain the lusory attitudes of the five player types of Salen and Zimmerman: Standard Player, Dedicated Player, Unsportsmanlike Player, Cheating Player, Spoil-Sport!
    - Standard : follows rules
    - Dedicated : follows rules but unusual strategies
    - Unsportsmanlike : follows rules but violates spirit of lusory attitude
    - Cheating : violates rules to win
    - Spoil-Sport : violates rules, doesn’t care at all

- Give your own definition of “Game Mechanics” considering the definitions by Hunicke, Järvinen, Sicart or Dormans!
    - “are the various **actions behaviors and control mechanisms** afforded to the player within a game context. together with the game’s **content(levels, assets and …)** the mechanics support overall gameplay dynamics”

- Briefly characterize Casual Games as one of four types of games discussed!
    - four types : Hardcore, Casual, Pervasive, Serious
    - Casual →
        - instant play → easy to learn
        - quick play → do not require much time to play to get pleasure
        - common play → address a vast majority of player types

- Name and briefly characterize two classes of Casual Games!
    - **meta-types** : online(primary), social(primary), mobile
    - **constant development** e.g. via user feedback possible and good practice

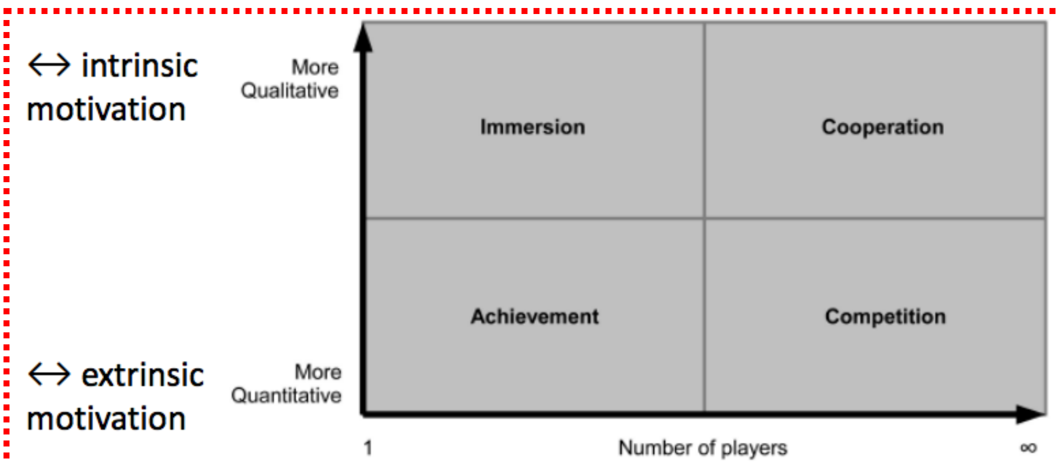
- Reasonably characterize the four types of games {Hardcore, Casual, Pervasive, Serious} in terms of the five meta-types of games {Simulation, On-Line, Social, Mobile, Location-based} in a table!

	Hardcore	Casual	Pervasive	Serious
Simulation	◕	○	○	◕
On-line	◔	◑	◔	◔
Social	◔	◑	◔	◔
Mobile	○	◔	◕	○
Location-based	○	○	◑	○

- What are advantages when characterizing game genres in terms of elements of game mechanics compared to characterizing game genres in terms of ludological genre frameworks?
    - genre classification is based on game mechanics: “genre is defined bt a shared collection of core mechanisms”
        - similar to ludological genre frameworks
        - **genres == set** of game mechanics
        - **additive** : new mechanics can be added without changing older parts
        - **new genres easy** : new combinations / sets

- Maslov’s Need Hierarchy contains the levels Physiological, Safety, Belonging-Love, Self-Esteem, Self-Actualization. Which levels do Games and Social Media contribute to? Give a brief explanation!
    - Self Esteem level
    - Belonging Love level (←→ social game)

- Briefly explain Radoff’s Player Motivations diagram!



    - social → more powerful motivators : acceptance or status

- Provide three examples of rewards systems and briefly explain their nature and motivational function and characterize them in terms of Wang and Sun’s four characteristics of reward (**social value, effect on game play, suitability for collection and review, time required to earn / receive the award**)!
    - score systems
    - experience point reward systems
    - item granting systems reward

Social Media

- Define “Social Media Service” and “Social Media Platform”!
  - Social Media Service : Web-based service, in the sense of SOA supporting (direct and indirect) social interaction (especially communication) via the generation and exchange of large amounts of content by a broad (compared to the number and nature of Internet users), non-IT-specialist set of users.
  - Social Media Platform : “functionally coherent bundle of Social Media services” (distinction service ↔ platform often not totally sharp) + commonly accessible, sufficiently widespread, distributed, functionally coherent bundle of network technologies on which it operates.

- **Characterization of Social Media services in terms of the supported forms of communication: Name and define four different characteristics (axes)!**
  - **Cardinality** of persons involved in a typical communication act **{1:1, 1:n, m:n}**
  - **Directedness** → **{direct, indirect}** → specific dedicated receive / list of receivers vs. open set of receiver, possibly formally or informally constrained e.g. via certain properties
  - **Anonymity**→ **{non-anonymous, anonymous}** → the identity of the sender(s) is or is not known to the receiver(s)
  - **Content**→ **{textual, graphical, video, contextual (** locations, social relations, user-item-relations etc.**)}**

- Social Semantic Web: What is the nature and function of OWL, SIOC and OPO?
  - OWL → semantic web languages
  - SIOC → semantically-interlinked online communities
    - enable the integration of online community information
    - provides a Semantic Web ontology for representing rich data from the Social Web in RDF
  - OPO → online presence ontology
    - enable the integration and exchange of Online Presence related data
    - representing rich data about online presence in RDF

- What are reasonable classes of Context in Social Networking and Mobile Social Networking? Briefly explain!
  - computing, user, physical, temporal
  - identity, location, status

- Social Networks: What information can be modeled in edge profiles? Three coarse classes of Social Media: briefly characterize Awareness (Contextual) services!
    - **inform** users about **events or states** directly linked with other users that fulfill certain(contextual) criteria, proactively or on request
    - **manage** contextual data (social network, privacy setting)
    - primary form of **content:** contextual information
    - typical **form of communication:** 1:n and m:n ; indirect ; non-anonymous ; non-threaded ; contextual
    - **example sub-class:** location-based awareness services

- What are main differences between Blogs and Microblogs?
    - supported typical **communication form**
        - **Blog** → 1:n, indirect, non-anonymous, non-threaded, textual (+ photos), desktop or laptop, discrete transfer, noncommercial
        - **Microblogs** → 1:n; indirect (and also direct), non-anonymous, non-threaded, short textual, mobile, discrete transfer, non-commercial

- Define “Social Game”!
    - commonly refers to playing games as a way of social interaction, as opposed to playing games in solitude.

- Briefly discuss the validity for Social Media of the following defining aspects of “play” or characterizations of some forms of “play”:
    - ‘outside ordinary life, not serious (Huizinga)’,
        - SM is not “outside ordinary life”, but rather outside many aspects of “serious” life . SM seen as communication support cross divisional role of SM
    - ‘unproductive’ (Caillios),
        - must be negated for knowledge codification, collaboration classes etc.
    - having ‘rules’ (Huizinga and others),
        - rules of social interaction, emergent special rules of SM (e.g. “nettiquette”)
    - ‘transformative social play’ (salen and Zimmerman)
        - many examples for transformative use in SM

- Name and briefly explain three common aspects of Social Games and Social Media!
  - mostly “outside” serious life, **leisure time oriented** → but both: increasingly many “**serious forms**”
  - **communication** as important element
  - defined set of **rules**
  - **emergent** mechanics & dynamics; transformative use
  - **transmedial** access patterns, blurring: real world ↔ virtual world
  - complex game worlds ↔ social information spaces
  - parallels in aspects of **motivation**, flow
  - one often the “**host**” of the other (e.g. in Social Networking games)

- Name and briefly explain three characteristics of Social Networking!
  - **Social Media services** → users socially interact using bundle of Social Media services (direct communication, information, awareness)
  - users have **Personal information spaces** → sets of items associated with users that they exert control over or whose relations (user-item) they exert control over
  - a user has **Personal profile:** publicly accessible sub-space of p.i.s.: used as personal reference: for introducing a person or used as reference point for SN services (e.g. awareness services)

- Define Mobile Social Networking!
  - comparable to social networking with an emphasis on mobile usage, contextual content elements and context awareness of services

Social Network Analysis → Centrality

- Directed graph with edge semantics (a) „a votes for b“ or (b) „a has convinced b“: What type of degree based centrality measure can be applied in case (a), what type in case (b)? (just note the two, no explanation required)
    - (a) „a votes for b“ → Winner “most central node” : node with most incoming edges (highest in-degree) → Degree Centrality
    - (b) „a has convinced b“ → Variant (influence network) → node with large out-degree is central

- Explain the expression *c(u) = 1/e(u) = 1/max{d(u,v):v belongs V}* for eccentricity-based centrality! (2-3 short sentences)
    - center of a graph → set of all nodes with minimum eccentricity
    - nodes in the center of the graph have maximal centrality

- Give the mathematical definition for closeness centrality and motivate it in one sentence referring to the Minisum problem!
    - find nodes whose sum of distances to other nodes is minimal (→ service facility location problem) : For all u minimize total sum of minimal distance $\sum_{v \in V} d(u,v)$

- **Using the following two graphs as examples, motivate an advantage of Shortest Path Betweenness centrality *c(v) = sum sum (delta_ab(v))* compared to simple Stress centrality *c(v) = sum sum (alpha_ab(v))***
    - Simple Stress centrality *c(v) = sum sum (alpha_ab(v))*
    - Shortest Path Betweenness centrality *c(v) = sum sum (delta_ab(v))*
        - Interpretation: Control that v exerts on the communication in the graph
        - **applicable to disconnected graphs**
        - variational ways regarding the path between a & b

- **Give a mathematical definition for general Vitality! What is the basic intuition behind the vitality type of centrality indices? (1 sentence)**
    - **Intuition**: Measure importance of vertex (or edge) by the difference of a given quality measure q on G with or without the vertex (edge):
        - Vitality v(x) of graph element **x : v(x) = q(G) - q(G\{x})**

- Give a reason to use flow-based centrality measures instead of shortest-paths-based centralities! (1 sentence)
    - major critique w.r.t. spb: resources (information, goods, work, rumors etc.) do not flow along shortest paths only.

- Newman's Random Walk Betweenness Centrality: We know that the transition matrix with column t removed is defined as *M_t = A_t * (D_t)^(-1)*. What is the probability that (starting at node s) we arrive after r steps in node j and then transition to node i immediately afterwards?
    - for a walk starting at s, the probability that we find ourselves at vertex j after r steps is given by $[\mathbf{M}_t^r]_{js}$
    - probability that we then take a step to an adjacent vertex i is $k_j^{-1}[\mathbf{M}_t^r]_{js}$.

- Derive an expression for the the total number of times V we hit the graph's nodes when doing a random walk starting at s and ending at t !

$$k_j^{-1}[(\mathbf{I}-\mathbf{M}_t)^{-1}]_{js}.$$

$$\rightarrow \quad \mathbf{V} = \mathbf{D}_t^{-1} \cdot (\mathbf{I}-\mathbf{M}_t)^{-1} \cdot \mathbf{s} = (\mathbf{D}_t - \mathbf{A}_t)^{-1} \cdot \mathbf{s}.$$

- **State the general idea of the family of feedback centrality measures! (1 sentence)**
    - page rank like indices
    - node is more central the more central its neighbors are

- Prove that for a random surfer model with adjacency matrix A and Markov transition matrix *t_ij = (A_ij)/(deg(i)) ⇒ (deg(i))/sum(deg(v))* is the stationary distribution!
    - Stationary distributions degree centrality: Assume undirected, unweighted graph with adjacency matrix A; we have then:

$$t_{ij} = \frac{A_{ij}}{\deg(i)} \Rightarrow \pi_i = \frac{\deg(i)}{\sum_{v \in V} \deg(v)}$$

Proof: $$(\pi T)_j = \sum_{i \in V} \pi_i t_{ij} = \frac{\sum_{i \in V} \deg(i) t_{ij}}{\sum_{v \in V} \deg(v)} = \frac{\sum_{i \in V} A_{ij}}{\sum_{v \in V} \deg(v)} = \frac{\deg(j)}{\sum_{v \in V} \deg(v)} = \pi_j$$

Social Network Analysis → Dense Subnetworks

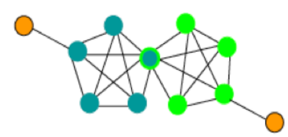
- In how far can Emergence as a point of view mediate between individualistic and collectivistic points of view when investigating groups? (2 -3 sentences)

- Nominate two characteristics of small groups!
  - number of group members <20
  - friends clique

- What are quasi groups? (1 sentence)
  - profile clusters only

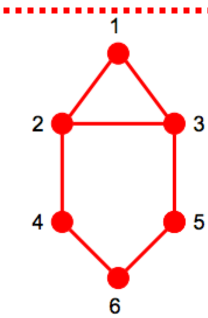
- Is a (Web-)community a group in the sociological sense? Give one supporting argument and one counter-argument!
  - Community : “A set of people which have a high degree of community-awareness, communicate with other members via electronic media, and have a common pursuit which can be identified with the pursuit to collaboratively build up a thematically focused, information or knowledge-space. This collaborative information- or knowledge-space (CIKS) predominantly contains semi-formal, implicit, “warm” information or knowledge with a strong emphasis on textual form”
  - Compare from community Platform characteristics & detail a More loosely defined compared to sociological definition; rather a “large group!

- **Can cliques overlap? Explain Your answer briefly!**
  - Yes! If Graph is dense à cliques exist
    → If $|E| > |V|^2/2\ (k-2)/(k-1)$ then G contains a clique of size k
  - Usually many different maximal cliques exist in a graph; they can overlap without being identical



- Given the following definitions for alternative prototypes
  - U is N-clique iff $\forall u,v \in V : distG(u,v) \leq N$
  - U is N-club iff $diam(G([U])) \leq N$
  - U is N-clan iff U is maximal N-clique and $diam(G([U])) \leq N$

  → Name all N-cliques, N-clubs, N-clans in the following graph!: p.17



maximal 2-cliques: {1, 2, 3, 4, 5}, {2, 3, 4, 5, 6}
2-clubs: {1, 2, 3, 4}, {1, 2, 3, 5}, {2, 3, 4, 5, 6}
2-clan: {2, 3, 4, 5, 6}

- Explain why N-cliques do not have to be connected!
    - Since dist is evaluated w.r.t. to G and not G([U]) (thus N-cliques are not local structures), N-cliques need not even be connected and can have a diameter diam(G([U]) > N
- Are N-clubs closed under exclusion and are they nested? Explain briefly!
    - Small distances are characteristic even for large social networks (cmp. 6 degrees) a N-cliques, N-clubs, and N-clans may **not be socially meaningful as groups** but may spheres (e.g. regarding information flows) → These constructs are **not generally closed under exclusion** and are **not nested** (socially meaning characteristics that cliques possess)

- **What are considerations when choosing N if N-cliques are used as prototypes for social groups in a social network? (1-2 sentences)**
    - small N is not meaningful
    - however its hard to find large N

- **Is the problem „Find the size k of a maximum clique of an undirected graph G" efficiently solvable? (1-2 sentences (informal explanation))**
    - **Probably not**: The decision problem CLIQUE(G,k): "Has G a clique of size at least k" is **NP complete.** (Solving the decision problem in time T(|V|) would yield an alg for determining the maximal k in O(T(|V|) log (|V|)) via binary search.

Social Network Analysis → Graph Clustering

- State the central paradigm for formulating a quality measure for clustering (target function for optimization)! (in words)
  - **Quality measure**: Objective function A(G) ℝ that formalizes the clustering paradigm in a special way
  - G= (V,E,w): **Weight function** w: E ℝ+ is interpreted as "**similarity**" (higher weights correspond to more intense tie); also possible: negative weights = dissimilarity; or w: E [0,1] or w: E [-1,1] etc.
  - **Distinguish** between no edge and edge with weight zero;
  - **Notation**: w(E) = ∑e∊Ew(e)

- Coverage is defined as $\gamma(\mathbf{C}) = \frac{w(E(\mathbf{C}))}{w(E)} = \frac{\sum_{e \in E(\mathbf{C})} w(e)}{\sum_{e \in E} w(e)}$. State a main disadvantage of using f(C)=γ(C) in a quality measure for clustering! (1 – 2 sentences)
  -

- Explain the pragmatics of the following two conductance based quality measures! (1-2 sentences each)

First measure: g=0 and $f(\mathbf{C}) = \min_{1 \le i \le k} \varphi(G[C\_i])$ Second measure: f=0 and

$$g(\mathbf{C}) = \begin{cases} 1 & \textit{if } \mathbf{C} = \{V\} \\ 1 - \max_{1 \le i \le k} \varphi(C\_i,\ V \setminus C\_i) & \textit{otherwise} \end{cases}$$

  - First measure:
    - <u>If the measure is small</u>: At least one of the clusters (more precisely: the induced subgraph) contains at least one bottleneck This cluster is too coarse Use minimum conductance cut to cut this cluster in "halves"
    - <u>From theorem before</u>: Only clusterings where the clusters induce subgraphs that are stars or have size at most three have f=1 (f is called intra cluster conductance)
  - Second measure:
    - <u>If the measure is small</u>: At least one of the clusters (more precisely: the induced subgraph) has many connections to outside The clustering is too fine Merge clusters
    - <u>From theorem before</u>: Only clusterings that have inter cluster edge weight zero have g=1 (g is called inter cluster conductance)

- Explain the pragmatics of the quality measure Performance! (2 sentences)

$$f(\mathbf{C}) = \sum_{i=1}^{k} |E(C\_i)|$$

$$g(\mathbf{C}) = \sum_{u,v \in V} [(u,v) \notin E] * [u \in C\_i, v \in C\_j, i \neq j]$$

  - Main idea: Clustering paradigm Count “correctly classified pairs of nodes”. A pair of nodes is correctly classified if:
    - It is in the same cluster AND connected by an edge f counts the number of edges within clusters
    - If it is not in the same cluster AND not connected by an edge g counts the number of non-existent edges between clusters

- In case a density measure π on graphs is available, a quality function can be constructed by pessimistically setting $\text{worst case: } \min_i \{\pi(G[C_1]), \ldots, \pi(G[C_k])\}$

  (a) **What are analogous expressions for average case and best case (optimistic view)?**

$$\text{average case: } \frac{1}{k} \sum_i \pi(G[C_i])$$

$$\text{best case: } \max_i \{\pi(G[C_1]), \ldots, \pi(G[C_k])\}$$

  (b) Suggest a simple density measure π on graphs!

  (c) **Explain in 1-2 sentences why finding π is easy if an embedding of the graph into a metric space is known!**

- Informally explain the difference between Linkage-based (Agglomerative) and Splitting-based (Divisive) hierarchical clustering! (1-2 sentences)
    - **Linkage (Agglomeration)**: Iteratively coarsens a given clustering by merging two clusters until 1-clustering is reached ("bottom up")
    - **Splitting (Division)**: Iteratively refines a given clustering by splitting one cluster until singleton clustering is reached ("top down").

- **Explain the difference between Complete Linkage and Single Linkage by stating the local cost functions for both cases and giving 1 sentence of explanation!**
    - Variants / realizations of Linkage:
        - Let d(u,v) denote the minimal path length between nodes u and v then local cost function:

$$c_{local}(C_i, C_j) = \substack{\max \\ \min}\{d(u,v) \mid u \in C_i, v \in C_j\}$$

Complete Linkage   Single Linkage

    - Variant of the Single Linkage and Complete Linkage Algorithm for weighted graphs (large weight == nodes are strongly related ("close")) using threshold graphs

- What is the benefit of a cut-function when using Splitting-based (Divisive) hierarchical clustering? (1-2 sentences)
    - cut function avoids having to test all possible splits
    -

- Newman – Girvan method: Explain Modularity! (2-3 sentence)

$$Q = \sum_i (e_{ii} - a_i^2) = \mathrm{Tr}\,\mathbf{e} - \|\mathbf{e}^2\|$$

    - intra cluster coherence (f) cluster validity measure (g=0) to optimally cut dendrogram
    - p.41

Data Mining → Metric Clustering

- Explain briefly the following three characterizations for clustering-approaches / methods:
  - exclusive vs. non-exclusive
    - Exclusive → non overlapping clusters
    - non-exclusive → overlapping clusters
  - crisp vs. fuzzy
    - Crisp clusterings → Conventional characteristic functions α_k for each Cluster C_k
    - Fuzzy clustering → fuzzy membership function α _k for each Cluster C_k (p.7)
  - hierarchical vs. non-hierarchical
    - Hierarchical clustering imposes a tree structure (Dendrogram) on the C_k where an edge C_i C'_j implies C_i ⊂ C'_j;

- The local cost functions for Single Linkage (min) and Complete Linkage (max) are defined as $c_{local}(C_i, C_j) = \begin{matrix}\max\\ \min\end{matrix} \{d(u,v) \mid u \in C_i, v \in C_j\}$ where d(u,v) denotes the length of the shortest path between nodes u and v. State the analogous local cost functions if nodes correspond to pattern vectors xu and xv in ℝn!
  - Completely analogous to graph clustering case: Start with singletons and on each level of the dendrogram merge two clusters with minimal distance (cost)

- Single link:

$$d(\mathcal{C}_{k_1}, \mathcal{C}_{k_2}) = \min_{\{n_1, n_2 | x_{n_1} \in \mathcal{C}_{k_1} \wedge x_{n_2} \in \mathcal{C}_{k_2}\}} ||x_{n_1} - x_{n_2}||$$

- Complete link:

$$d(\mathcal{C}_{k_1}, \mathcal{C}_{k_2}) = \max_{\{n_1, n_2 | x_{n_1} \in \mathcal{C}_{k_1} \wedge x_{n_2} \in \mathcal{C}_{k_2}\}} ||x_{n_1} - x_{n_2}||$$

- K-means clustering:
  - (a) Define an objective function (cluster quality index) for K-means!
    - Optimize intra cluster coherence:
    - Describe cluster C_k by prototype μ_k; prototype need not be an actual pattern (If so, algorithm works with slight modifications as well)
    - Determine cluster for each pattern x_n by nearest neighbour rule:

$$\boxed{\mathscr{C}(x_n) = k_a \leftrightarrow ||x_n - \mu_{k_a}|| = \min_i ||x_n - \mu_k||}$$

  - (b) What is the resulting simple expression for the prototypes μk?
    - Find prototypes by optimizing objective function modeling intra cluster coherence as mean square error

$$J_{\mathrm{SQE}} = \sum_{k=1}^{K} \sum_{\{n|x_n \in \mathcal{C}_k\}} ||x_n - \mu_k||^2$$

$$\frac{\mathrm{d}J_{\mathrm{SQE}}}{\mathrm{d}\mu_k} \overset{!}{=} 0 \implies \mu^k = \frac{1}{|\mathcal{C}_k|} \sum_{\{n|x_n \in \mathcal{C}_k\}} x_n$$

    → cluster prototypes are barycenters („centers of gravity“) of their clusters

- State and explain two advantages of DBSCAN (compared to K-Means)! (1 sentence each)
  - favors spherical clusters → arbitrarily shaped clusters
  - need to know K → We do not need to know K in advance
  - no notion of noise → notion of noise

- Fuzzy C-Means:
  - What is the range of values of the entries membership matrix r_nk compared to K-means? p.18
  - **Fuzzyness-parameter m in the objective function: Informally: what happens in the limits m → 1 and m → ∞ ? (no full mathematical derivation necessary!)**
    - m → 1 : K-Means (crisp case);
    - m → ∞ : rnk 1/K (where K is the number of clusters)

  - Explain the pragmatics behind the additional conditions

$$\forall x_n \;:\; \sum_{k=1}^{K} \alpha_k(x_n) = \sum_{k=1}^{K} r_{nk} = 1$$

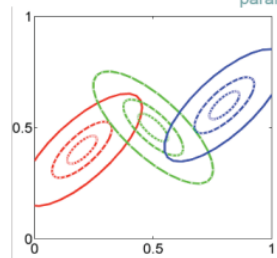
$$\forall \mathcal{C}_k \;:\; \sum_{n=1}^{N} \alpha_k(x_n) = \sum_{n=1}^{N} r_{nk} > 0$$

optimization for the object function!

- **What are advantages of Gaussian Mixture Models compared to Fuzzy C-Means? Informally explain the nature and geometrical interpretation of the GMM quantities μk and Σk!**

$$p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k) \quad \text{where} \quad \sum_{k=1}^{K} \pi_k = 1, \quad 0 \leqslant \pi_k \leqslant 1$$

parameters to be estimated



[6]

- What is the paradigm of the „Maximum Likelihood concept“?
  - finding the Θ that best explains the data
  - Maximum Likelihood: ΘML = $argmax\Theta\ p\ X\ \Theta$ ⇨ $\nabla\Theta\ p\ X\ \Theta$ = 0

- What does „iid“ mean? What are the consequences here? Can You point them out in an expression on these slides?
  - •iid: „identically independently drawn“ ⇨ $p\ X\ \Theta$ = $i\ p(xi$ |Θ)
    - p X Θ is called likelihood
    - „finding the Θ that best explains the data“: Maximum Likelihood: ΘML = $argmax\Theta\ p\ X\ \Theta$ ⇨ $\nabla\Theta\ p\ X\ \Theta$ = 0
    - convenient: use log $p\ X\ \theta$ instead of $p\ X\ \theta$ ⇨ log $p\ X\ \Theta$ = i log $p(x\_i$|Θ)

- **Why do we take the logarithm of the likelihood? How do we choose the base of the logarithm ? Why?**
  - convenient → use log $p(X|\theta)$ instead of $p(X|\theta)$ ⇨ log $p(X|\Theta)$ = i log $p(x\_i$|Θ)

• log likelihood: (use base e)

$$\ln\ p(\mathbf{X}|\Theta) = \ln p(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = -\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{n=1}^{N}(\mathbf{x}_n - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}_n - \boldsymbol{\mu})$$

• Maximum log likelihood:

$$\Theta_{ML} = argmax_{\Theta}\ \log p(X|\Theta) \Rightarrow \nabla_{\Theta}\left(\sum_i \log p(x_i|\Theta)\right) \overset{!}{=} 0$$

$$\left.\begin{aligned} \boldsymbol{\mu}_{\mathrm{ML}}:&\ \frac{\partial}{\partial \boldsymbol{\mu}} \ln p(\mathbf{X}\,|\,\mu, \boldsymbol{\Sigma}) = 0 \\ \boldsymbol{\Sigma}_{\mathrm{ML}}:&\ \frac{\partial}{\partial \boldsymbol{\Sigma}} \ln p(\mathbf{X}\,|\,\boldsymbol{\mu}, \boldsymbol{\Sigma}) = 0 \end{aligned}\right\} \Rightarrow \begin{aligned} \boldsymbol{\mu}_{\mathrm{ML}} &= \frac{1}{N}\sum^{N}\mathbf{x}_n \\ \boldsymbol{\Sigma}_{\mathrm{ML}} &= \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \boldsymbol{\mu}_{\mathrm{ML}})(\mathbf{x}_n - \boldsymbol{\mu}_{\mathrm{ML}})^{\mathrm{T}} \end{aligned}$$

- Case ONE Gaussian: Are the resulting equations directly solvable?

- What is a 1 of K representation?

K-dimensional binary random variable $\mathbf{z}$

$z_k \in \{0,1\}$ and $\sum_k z_k = 1$

$p(z_k = 1) = \pi_k$

$$p(\mathbf{z}) = \prod_{k=1}^{K} \pi_k^{z_k}$$

- What is the meaning of the latent variables Z = {znk}?
  - variables that are not directly observed but are rather inferred from other variables that are observed (directly measured)
  - determine the component from which the observation originates

  - if we have several observation x1,....xn, then, because we have represented the marginal distribution in the form p(x) = sum_z p(x,z), it follows that for every observed data point xn there is a corresponding latent variable zn

- Can You explain the formula for the responsibilities from an intuitive point of view?

$$\begin{aligned}\gamma(z_k) \equiv p(z_k = 1|\mathbf{x}) &= \frac{p(z_k = 1)p(\mathbf{x}|z_k = 1)}{\sum_{j=1}^{K} p(z_j = 1)p(\mathbf{x}|z_j = 1)} \\ &= \frac{\pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}.\end{aligned}$$

- Why don‘t we have the full joint probability distribution p(X,Z|θ)?
  - if we have the complete dataset {X, Z}, and thus the distribution p(X,Z|θ), we could use ML to solve for θ with p(X,Z|θ) directly, which is easy, as well we see, because p(X,Z|θ) is of exponential family (the functional form is known)
  - we only know p(X,Z|θ) (→ responsibilities, as well we see) → compute expectation of (unknown) quality p(X,Z|θ) or even better of the quality ln p(X,Z|θ)

- In general EM: Why do we take the expectation of the quantity ln p(X,Z|θ) under the probability distribution p(Z|X,θ old) ? Can You explain the formula for Q(θ,θ old) intuitively?

Real World Networks: Properties and Models

- Informally define the term „Small World Effect“! (1 sentence).
  - l(n) “small” → l(n)ϵ O(log(n))

- What is the disadvantage of the expression $\ell = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij}$ for the average path length in a network? (1 sentence). Suggest an alternative expression avoiding this disadvantage!
  - this expression only works for connected nodes
  - expression for disconnected points → harmonic mean

- State the expression for the fraction of nodes with degree k (==the probability of a node having degree k) in case of a power law degree distribution!

  - Notation:

    $p(k) = p_k =$ fraction of nodes having degree k

  - Cumulative distribution:

    $$P_k = \sum_{k'=k}^{\infty} p_{k'}$$

  - power law:

    $$p_k \sim k^{-\alpha}$$

    $$\rightarrow \quad P_k \sim \sum_{k'=k}^{\infty} k'^{-\alpha} \sim k^{-(\alpha-1)}$$

  - exponential:

    $$p_k \sim \mathrm{e}^{-k/\kappa}$$

    $$\rightarrow \quad P_k = \sum_{k'=k}^{\infty} p_k \sim \sum_{k'=k}^{\infty} \mathrm{e}^{-k'/\kappa} \sim \mathrm{e}^{-k/\kappa}$$

- If a power law distribution has an exponent of (-α): What is the exponent of the corresponding cumulative distribution?

$$\boxed{p_k \sim k^{-\alpha} \quad \rightarrow \quad P_k \sim \sum_{k'=k}^{\infty} k'^{-\alpha} \sim k^{-(\alpha-1)}}$$

- **Sketch a power law distribution in a log-log coordinate system!**
  - a straight line!

- What are the two disadvantages for the following measure of Assortativity: $Q = \frac{\sum_i P(i|i) - 1}{N-1}$

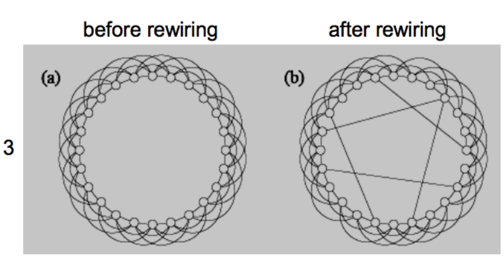
  with $P(j|i) = e_{ij} / \sum_j e_{ij}$ and a normalized mixing matrix $\mathrm{e} = \frac{\mathbf{E}}{\|\mathbf{E}\|}$; $\sum_{ij} e_{ij} = 1$
  - Asymmetry of E → two values
  - Not respecting size of classes

- Give a precise expression for the degree distribution pk in a Poisson Random Graph model G_n,p ! Qualitatively sketch pk for n → ∞, keeping the mean degree z fixed ! (Hint: the Poisson approximation for p_k is p_k ~)
  - G_n,p : space or graphs with **n-nodes** and each of the 1/2n(n-1) **edges appears with probability p**
  - p_k : probability that a node has degree k: $\boxed{p_k = \binom{n}{k} p^k (1-p)^{n-k} \simeq \frac{z^k \mathrm{e}^{-z}}{k!}}$ → for **n→ infinity** and holding the **mean degree of a node z = p(n-1) fixed** **(**Poisson approximation of Binomial distribution**)**

- Emergence of a giant component in a Poisson random graph model: Explain the following expression for the fraction of nodes u that do not belong to the giant component:

- Watts-Strogatz model: informally describe the model for D=1!
  - L nodes in regular D-dim. lattice + periodic boundary cond.; **D=1: Ring**
  - each node connected to neighbors in lattice at distance of most k → total number of edges = L k
  - „rewiring“ of edges with probability p

D = 1; k = 3



- Watts-Strogatz model: Interpret / explain the following diagram in view of the rewiring parameter p! (1-2 sentences)
  - rewire both „ends“ of edges + allow self-edges +.... math.easier
  - only add additional shortcut edges (no rewiring)
    - mean total number of shortcuts = L k p
    - mean degree of each node = 2k(1+p)

- Watts Strogatz model, variant (2) (only additional shortcuts) : degree distribution:

$$p_j = \binom{L}{j-2k}\left[\frac{2kp}{L}\right]^{j-2k}\left[1-\frac{2kp}{L}\right]^{L-j+2k}$$

for $j \geq 2k$, and $p_j = 0$ for $j < 2k$.

  - Why is pj = 0 for j < 2k?
  - What is the meaning of the expression 2kp/L ? → **Binomial distribution**

- Model of Price: Informally motivate the terms on the right side of the recursion equation for the net change of npk per added vertex:

$$(n+1)p_{k,n+1} - np_{k,n} = \left[kp_{k-1,n} - (k+1)p_{k,n}\right]\frac{m}{m+1}$$

  - mean number of nodes with in-degree k (which is npk) decreases by x because their in-degree changes to k+1
  - mean number of nodes with in-degree k also increases because of nodes having previously k-1 and now have k

  → the net change in the quantity npk per added vertex satisfies:

$$(n+1)p_{k,n+1} - np_{k,n} = \left[kp_{k-1,n} - (k+1)p_{k,n}\right]\frac{m}{m+1}$$

for $k \geq 1$, or

$$(n+1)p_{0,n+1} - np_{0,n} = 1 - p_{0,n}\frac{m}{m+1},$$

for $k = 0$.

- Network resilience: what is the difference in effect between randomly removing nodes with (a) constant probability q or (b) removing the highest degree nodes? (Explain informally in 1-2 sentences!
  - For power law networks:
    - (a) → remove random nodes : no effect on mean distances
    - (b) → remove high degree nodes: drastic effect

- Models of Epidemiology: quantities: susceptibles s, infectives i, recovered r, infection rate β, recovery rate γ: Explain the pragmatics of / interpret the following set of differential equations: $\frac{\mathrm{d}s}{\mathrm{d}t} = -\beta i s, \qquad \frac{\mathrm{d}i}{\mathrm{d}t} = \beta i s - \gamma i, \qquad \frac{\mathrm{d}r}{\mathrm{d}t} = \gamma i$
    - <u>susceptibles</u>: can be infected;
      <u>infective</u>: have the disease and are contagious,
      <u>recovered</u>: have had the disease and are immune (or dead)
        - infection probability / rate β , recovering probability γ
        - SIR model („fully mixed“)

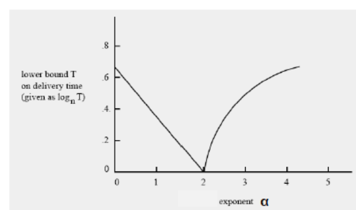
Spatial Properties of Social Relationship

- Describe the function of the parameters q1 and q2 and α in the Kleinberg model! (1 sentence per parameter)
  - q1 → the maximum distance of node connect locally
  - q2 → the amount of “long range” edge connection

- **Define„local knowledge“ in the context of a „greedy network routing algorithm with only local knowledge“ (e.g. in connection with the Kleinberg model)**
  - Each node only knows only:
    - its adjacent nodes
    - the grid’s principle structure
    - position of target node on the grid
    - positions and long-range contacts of nodes on the message path so far

- In a two-dimensional Kleinberg model the following statements concerning the expected number of delivery steps are true:
  - 0 ≤ α < 2 : s at least ~ c1(α,q1,q2) n(2-α)/3
  - α = 2 : s at most ~ c2(α,q1,q2) (log n)2
  - α > 2 : s at least ~ c3(α q1 n(α-2)/(α-1)

  For which value of α is efficient delivery possible? Give a semi-formal explanation for this behavior! (1-2 sentences)
  - α = 2 → is necessary for efficient delivery
  - Further analysis : if grid has dimension D, α = D is necessary for efficient delivery
  - Especially : if α = 0 (uniform probability for long range contacts as in Watts Strogatz model : no efficient delivery possible)



- Fora 2-dimensional Kleinberg model, α=2 is required for efficient delivery with greedy routing with only local knowledge. Liben-Nowell‘s finding from 2005 investigating greedy routing with only local knowledge on a real social network found that α ≈ 1 and efficient delivery is never the less possible. What is a possible explanation? (1 short sentence)
  p.11

Social Signal Processing

- Define „Social Intelligence“ for IT systems! (1 sentence) which parts of Your definition apply to the field of Multi Agent Systems and which parts are related to Social Signal Processing?
    - Ability to express and recognize social signals / social behaviors from other human and IT-agent individuals in order to „function“ in a society with other human and IT-agent individuals in view of (pareto-) optimizing own and other IT agent‘s and fellow human‘s utility function (survival, reproduction, ...) via cooperation
        - human → social signal processing for useful services
        - IT-agent → multi-agent-systems

- Characterize Reality Mining! (1 sentence) What is the relation between Reality Mining and Social Signal Processing? (1 sentence)
    - Social signal processing → individual behavior model
    - Reality mining
        - Analyzing all available traces of humans behavior (social & non-social)
        - reality mining may use social signal process techniques

- Name 3 examples for social signals / social behavior and name 3 examples for behavioral cues
    - Social behavior : expressing attitude towards elements of a social setting
        - Mirroring (if mutual attraction)
        - aggressive turn taking behavior
        - expression disapproval of sth. (e.g. via disapproving looks)
        - expression of sympathy / empathy
    - Behavioral Cues
        - facial expressions
        - body posture / interaction geometry
        - gestures
        - expressives (laughter etc.)
        - emotions reflected in speech prosody (rhythm, intonation, stress)

- Define behavioral cue! (1 sentence) What is the relation between social signals and behavioral cues? (1 sentence)
    - **Behavioral cue** → „composed of“ / manifested via (series of / parallel / overlapping / single ...) time-series of perceivable or measurable, non-verbal physiological activity. (neglecting content of communication)
    - **Social Signals** (conscious or unconscious): „composed of“ / manifested via (series of / parallel / overlapping / single ...) Behavioral Cues

- What is prosody? (1 sentence)

    - Voice quality: (e.g. pitch, tempo, energy)
        - anger, fear → energy peaks
        - pitch → perception of dominance or extroversion
        - pitch accents, changes in energy → structure or emphasize sth.

- For SSP: What is the advantage of unconscious social signals vs. conscious social signals? (1 sentence)

- **Facial expressions: What are Action Units (AUs)? (1 sentence)**
    - Smallest discernable **movements of a distinct muscle** in a **face** which may take part in **facial expressions** and **facial actions** and that may be **algorithmically detected** in view of e.g. detecting and classifying facial expressions and actions.

- Name the 6 basic emotions (after Ekman)
    - fear, sadness, happiness, anger, disgust, surprise

- Vocal Behavior: What are Linguistic Vocalizations and Non-Linguistic Vocalizations? (For each: 1 sentence plus 1 example) What is Backchanneling? (1 sentence)
    - **Linguistic vocalizations (Segregates)**: “non-words”: e.g. “ah”, “äh”, “umh”, etc.: examples:
        - prolonged “äääähm” –> embarrassment / feeling uncomfortable in social situation
        - **backchanneling** (attention, agreement, wonder etc.)
    - **Non-linguistic vocalizations**: other verbal sounds: e.g. laughter, crying, groaning : examples:
        - used as social signals to express boredom, sexual interest, anxiety etc.

- **Vocal behavior: Name and explain in 1 short sentence each three classes of silence!**
    - Silence patterns:
        - **hesitiation silence: (e.g. explaining difficult concepts)**
        - **psycholinguistic silence: (language) en-/de-coding difficulties**
        - **interactive silence: expressing respect, doubt, ignoring persons, attract attention to other forms of communication (e.g. gazes)**

- Name and explain in 1 sentence each 3 steps / sub-problems of Speaker Diarization!
    - “Speaker Diarization / Segmentation“: given multi-party audio data (possibly with background noise):
    - Typically 3 steps:
        - **segmentation into speech / non-speech** → Use (several) trained binary classifier(s) to distinguish between speech and non-speech on the computed features
        - **detection of speaker transitions** → split the speech parts into segments (length: e.g. 2-3 seconds)
        - **clustering of speaker segments** (+ classification of speaker) → -- e.g. use hierarchical bottom up clustering: merge segments with most similar models (e.g. Gaussians); cut dendrogram at maximum total likelihood

- Coarsely define optical flow and derive the optical flow equation!
    - motion pattern of picture elements (e.g. pixels): represented by vector field of velocity V(x,y,t) of intensity:

$$I(x+dx, y+dy, t+dt) = I(x,y,t) + \frac{\partial I}{\partial x}dx + \frac{\partial I}{\partial y}dy + \frac{\partial I}{\partial t}dt + O(d^2$$

$$\longrightarrow \quad \frac{\partial I}{\partial x}V_x + \frac{\partial I}{\partial y}V_y + \frac{\partial I}{\partial t} = 0$$ (optical flow equation)

use numeric methods to compute solutions

- What is the role of context in Social Signal Processing?
    - the way to use SSP techniques for improving services
        - smart home:
            - adapt environment to **short term context** → **individual emotional state**
        - availability management:
            - use **short term social context** detected/ characterized via SSP : social situation → modify volume of ringtone

<u>Basic of Game Theory</u>

- Prove via iterated strict dominance, that for the Prisoner's Dilemma (D,D) is a (pure, strict) Nash-Equilibrium!

	C	D
C	1, 1	-1, 2
D	2, -1	0, 0

- What is the rational behavior in an infinitely repeated Prisoner's Dilemma? Give a short reasoning!
    - rational → cooperation (infinitely repeated)

- Assuming a set of players I with |I| = I and a finite space S = xi Si of pure strategy profiles s ∈ S. Derive the mathematical expression for the utility of player i ∈ I if a mixed strategy profile σ =(σ1, σ2,…, σI) is played! $\boxed{\sum_{s \in S} \left(\prod_{j=1}^{I} \sigma_j(s_j) \right) u_i(s)}$

- Explain why a pure strategy may be dominated by a mixed strategy even if it is not strictly dominated by any pure strategy, using the following example game:

	L	R
U	2, 0	-1, 0
M	0, 0	0, 0
D	-1, 0	2, 0

    - Player 1: M not dominated by U and M not dominated by D
    - But: If Player 1 plays σ1 =(1/2, 0, 1/2) he will get u(σ1)=1/2 regardless how player 2 plays

- Prove that a mixed strategy may be dominated even if it assigns positive probabilities to pure strategies that are not even weakly dominated, using the following example

	L	R
U	1, 3	-2, 0
M	-2, 0	1, 3
D	0, 1	0, 1

    - U and M are not dominated by D for player 1
    - But: Playing σ1=(½, ½, 0) gives expected utility u1 (σ1, *) = - 1/2 no matter what 2 plays D (σD =(0, 0, 1)) dominates σ1

- **What is a Vickrey Auction? Name the rational strategy in a Vickrey auction (no explanation required)!**
    - Good's valuations: vi ; Assume common knowledge for the moment
    - Bids: si
    - Second price:
        - winning condition: si > max j≠i sj
        - let ri := max j≠i sj   ri is the price having to be paid
        - winner i ´s utility: ui = vi – ri ; other players utility = 0

→ the highest bidder wins the good but the price paid is the second-highest bid.

    - rational strategy : bidding true value

- In a mixed strategy Nash Equilibrium: Why must a player be (a priori) indifferent between all pure strategies to which he assigns positive probability (not regarding the other player‘s choice)? (1 sentence)
    - because the expected utilities are “linear in the probabilities”

For player I’s utility, we have:

$$u_i(\sigma) = \sum_{s_i \in S_u} \sigma_i(s_i) u_i(s_i, \sigma_{-i}) \quad \text{with} \quad \sum_{s_i \in S_u} \sigma_i(s_i) = 1$$

for the NE σ^* we thus have:

$$u_i(\sigma^*) = \sum_{s_i \in S_u} \sigma_i^*(s_i) u_i(s_i, \sigma^*_{-i}) \quad \text{with} \quad \sum_{s_i \in S_u} \sigma_i^*(s_i) = 1$$

since $u_i(\sigma^*)$ is the best outcome , i can achieve, when the others play σ^*_{-i} , all the $u_i(s_i, \sigma^*_{-i})$ with $\sigma_i(s_i) > 0$ must be equal,
and equal to $u_i(\sigma^*)$.

why? → no $u_i(s_i, \sigma^*_{-i})$ can be greater than $u_i(\sigma^*)$ otherwise the NE condition would be violated, and also not smaller, because then the sum would also be smaller.

- **assume that in a 2 player game the mixed strategy profile ((a,b,0),(c,d,0)) is a mixed strategy NE. Does the Indifference Condition in a mixed strategy NE imply that a = b = ½ ? Give a short reasoning!**
    -

- In the Cournot competition, each firm i 's profit is given by ui (q1, q2) = qi p(q) - ci (qi). Derive the defining equation for the reaction function r2 (.) !

  • Under certain reasonable assumptions (see [1]) we can maximize e.g. $u_2(q_1, q_2)$ by solving $d/dq_2\ u_2(q_1, q_2) = 0$ which yields

  $d/dq_2\ [q_2\, p(q_1, q_2) - c_2(q_2)] = p(q_1, q_2) + p'(q_1, q_2)\, q_2 - c_2'(q_2) = 0.$

  Inserting $r_2(q_1)$ for q_2

  $p(q_1 + r_2(q_1)) + p'(q_1 + r_2(q_1))\, r_2(q_1) - c_2'(r_2(q_1)) = 0$

  gives the defining equation for $r_2(.)$.

  (analogous for $r_1(.)$).

  • The intersections of the functions r_2 and r_1 are the NE of the Cournot game.

  • Example: Linear demand $p(q) = \max(0, 1-q)$; linear cost: $c_i(q_i) = c\, q_i$:

  → $r_2(q_1) = 1/2\, (1 - q_1 - c)$; $r_1(q_2) = 1/2\, (1 - q_2 - c)$;

  → NE: $q^*_2 = r_2(q^*_1) = 1/3\, (1-c) = q^*_1 = r_1(q^*_2)$

- Explain, why ((1/2, 1/2); (1/2, 1/2)) is a mixed strategy NE in the Matching Pennies game!

	H	T
H	1, -1	-1, 1
T	-1, 1	1, -1

p.53

  - If player 2 plays (1/2, 1/2) then player 1's expected payoff is ½ *1 + ½ *(-1) = 0 when playing H and ½ *(-1) + ½ *1 = 0 when playing T player 1 is also indifferent

- **Derive the mixed strategy NE in the battle of the sexes, using the Indifference Condition!**

	B	F
F	0, 0	2, 1
B	1, 2	0, 0

  - One mixed NE: Indifference condition Let σ1(F)=x and σ2(B)=y
  - Player 1's indifference: 0 y + 2(1-y) = 1 y + 0 (1-y) y=2/3
  - Player 2's indifference: 0 x + 2(1-x) = 1 x + 0 (1-x) x=⅔
    → Mixed NE: ((2/3, 1/3); (2/3, 1/3))