V. GENERALISED RANDOM GRAPHS

ER graph has many properties that make it unsuitable as model for real word networks (degree distribution, homogeneity, etc). In the following 3 chapters we study various other models of random graphs that behave better in this direction.

The idea of the first model, called "generalised random graph", is to equip the vertices with weights that introduce some inhomogeneity in the model.

We thus consider n vertices and a weight function $w : [n] \to (0, \infty)$. We connect two vertices $i \neq j$ with probability

(5.1) $$p_{ij} = \frac{w_i w_j}{\ell + w_i w_j},$$

(5.2) where $\ell = \ell_n = \sum_{i \in [n]} w_i$ is the total weight

(5.3) <u>Exercise</u>: Show that if $w_i = \frac{n\lambda}{n-\lambda}$, then we obtain $ER(n, \frac{\lambda}{n})$

The weights w_i might be deterministic, or themselves random. Their possible dependence on n is kept implicit.

(5.4) <u>Example</u>: (Population of two types) $n = n_1 + n_2$

Assume we have $w_1 = \quad = w_{n_1} = m_1$, $w_{n_1+1} = \ldots = w_n = m_2$, i.e. there are two types of vertices. Then $\ell = n_1 m_1 + n_2 m_2$

Denoting $p^{\alpha\beta}$, $\alpha, \beta = 1, 2$, the probability that type α is conn. to type β, we have

(5.5) $$p^{\alpha\beta} = \frac{m_\alpha m_\beta}{\ell + m_\alpha m_\beta}$$

Hence, the expected degree of vertex of type 1 is

$$(n-1)\cdot\frac{m_1^2}{\ell+m_1^2}+n_2\,\frac{m_1 m_2}{\ell+m_1 m_2}=m_1\left[\frac{(n_1-1)m_1}{\ell+m_1}+\frac{n_2 m_2}{\ell+m_2}\right]$$
$$=m_1\,(1+o(1))$$

whenever $m_1^2+m_1 m_2=o(\ell_n)$.

Let F_n be the empirical dist. function of the weights

(5.6) $$F_n(a)=\frac{1}{n}\sum_{i\in[n]}\mathbb{1}_{\{w_i\le a\}}.$$

It can be also interpreted as dist. function of the weight of randomly selected vertex. Formally, let

(5.7) U be uniform on $[n]$, $W=w_U$. Then F_n is dist. function of $W=W_n$.

We want to understand degrees in GRG. We need certain regularity assumptions on the weights.

(5.8) <u>Condition</u>: There exists a distribution function F s.t.

(a) $W_n \xrightarrow[n\to\infty]{d} W$, where W has d.f. F.

(b) $E[W_n]\xrightarrow[n\to\infty]{} E[W]>0$

(c) $E[W_n^2]\xrightarrow[n\to\infty]{} E[W^2]$

(5.9) <u>Exercise</u>: (iid. weights) Assume W is a given $(0,\infty)$-valued r.v. with d.f. F, s.t. $EW\in(0,\infty)$, $EW^2<\infty$. Let $(w_i, i\in[n])$ be iid, with d.f. F. Show that the condition (5.8) is a.s. satisfied for a given realisation of $(w_i)_{i\in[n]}$.

We now explore some properties of GRG. We write $GRG_n(w)$ for the GRG with given n and w, and $E_{n,w}$ for its number of edges

(5.10) Theorem: Assume (5.8 a,b). Then

$$\frac{1}{n} E_{n,w} \xrightarrow{\mathbb{P}} \frac{1}{2} \mathbb{E}[W].$$

Proof: We apply the second moment method. Note that

$$\mathbb{E}[E_{n,w}] = \frac{1}{2} \sum_{i\neq j} p_{ij} = \frac{1}{2} \sum_{i\neq j} \frac{w_i w_j}{l_n + w_i w_j} \leq$$

(5.11)

$$\leq \frac{1}{2} \sum_{i,j} \frac{w_i w_j}{l_n} = l_n/2.$$

A corresponding lower bound can be difficult if $w_i w_j$ is large. We thus truncate. Fix $a_n \nearrow \infty$. Since $x \mapsto \frac{x}{l+x}$ is increasing,

(5.12)
$$\mathbb{E}[E_{n,w}] \geq \frac{1}{2} \sum_{i\neq j} \frac{(w_i \wedge a_n)(w_j \wedge a_n)}{l_n + (w_i \wedge a_n)(w_j \wedge a_n)}.$$

Denoting $\overline{w_i} = w_i \wedge a_n$, $\overline{l_n} = \sum_{i\in[n]} \overline{w_i}$, we have

$$\frac{\overline{l_n}^2}{l_n} - 2\mathbb{E}[E_{n,w}] \leq \sum_i \frac{\overline{w_i}^2}{l_n + \overline{w_i}^2} + \sum_{i,j\in[n]} \overline{w_i}\,\overline{w_j} \left[\frac{1}{l_n} - \frac{1}{l_n + \overline{w_i}\,\overline{w_j}}\right]$$

(5.13)
$$= \sum_i \frac{\overline{w_i}^2}{l_n + \overline{w_i}^2} + \sum_{i,j} \frac{\overline{w_i}^2 \overline{w_j}^2}{l_n (l_n + \overline{w_i}\,\overline{w_j})}$$

$$\leq \sum_i \frac{\overline{w_i}^2}{l_n} \left(1 + \sum_j \frac{\overline{w_j}^2}{l_n}\right)$$

It is easy to see that

$$\sum \frac{\overline{w_i}^2}{l_n} \leq a_n$$

We thus pick $a_n = o(\sqrt{n})$. Then the RHS of (5.13) is $o(n)$. We claim

(5.14)
(a) $$\frac{\overline{l_n}^2}{n\, l_n} \xrightarrow[n\to\infty]{} \mathbb{E}[W]$$
(b). $$\frac{1}{n} l_n \xrightarrow[n\to\infty]{} \mathbb{E}[W]$$

Then, (5.11) – (5.14) imply

(5.15)
$$\frac{1}{n} \mathbb{E}[E_{n,w}] \geq \frac{1}{2} \mathbb{E}[W](1+o(1))$$

To show (5.14), observe that (5.8(b)) implies (5.14(b)). Further we claim that $\frac{\overline{l_n}}{n} \longrightarrow \mathbb{E}[W]$. Indeed, $\limsup \frac{\overline{l_n}}{n} \leq \mathbb{E}W$ is trivial, and $\liminf \frac{\overline{l_n}}{n} \geq \mathbb{E}W$ follows by MCT (Exercise).

Bounding the variance is much shorter:

$$Var(E_{n,w}) = \frac{1}{2} \sum_{i \neq j} Var(X_{ij})$$

where X_{ij} are independent, $X_{ij} \sim$ Bernoulli(p_{ij}). Hence

(5.16) $$Var(E_{n,w}) = \frac{1}{2} \sum_{i \neq j} p_{ij}(1-p_{ij}) \leq \frac{1}{2} \sum_{i \neq j} p_{ij} = \mathbb{E}[E_{n,w}].$$

(5.10) then follows by the standard Chebyshev argument. $\square$

We now control the behaviour of degrees of GRG. Let

(5.17) $$X_{ij} = \mathbb{1}\{\text{edge } i,j \text{ is present in GRG}\} \sim \text{Bernoulli}(p_{ij}), \ i<j.$$
$$D_i = \sum_{j \neq i} X_{ij} = \text{degree of } i \text{ in GRG}.$$

(5.18) <u>Theorem:</u> (a) There is a coupling $(\hat{D}_i, \hat{Z}_i)$ of D_i with $Z_i \sim$ Pois(w_i), such that

$$\mathbb{P}(\hat{Z}_i \neq \hat{D}_i) \leq \frac{w_i^2}{\ell_n}\left(1 + 2\frac{\mathbb{E}W_n^2}{\mathbb{E}W_n}\right)$$

(b) If (5.8(a,b)) hold and $\lim_{n\to\infty} \max_{i,j \leq n} p_{ij} = 0$, then the degrees $D_1, \dots, D_m$ are ass. independent.

<u>Proof</u> (a) By Question 3 of Serie 1 of exercises D_i can be coupled to $Y_i \sim$ Pois(λ_i) with $\lambda_i = \sum_{j: j\neq i} p_{ij}$ with error

$$\mathbb{P}(\hat{D}_i \neq \hat{Y}_i) \leq \sum_j p_{ij}^2 \leq w_i^2 \sum_j \frac{w_j^2}{\ell_n^2} = \frac{w_i^2}{\ell_n} \frac{\mathbb{E}W_n^2}{\mathbb{E}W_n}$$

We then need to show that Y_i can be coupled to Z_i so that

(5.19) $$\mathbb{P}(\hat{Y}_i \neq \hat{Z}_i) \leq \frac{w_i^2}{\ell_n}\left(1 + \frac{\mathbb{E}W_n^2}{\mathbb{E}W_n}\right).$$

To this one observes that

$$\lambda_i = \sum_{j \neq i} \frac{w_i w_j}{\ell_n} \leq w_i \sum_j \frac{w_j}{\ell_n} = w_i$$

So, introducing $\hat{V}_i =$ Pois$(w_i - \lambda_i)$, $\hat{Y}_i \sim$ Pois(λ_i), $\hat{Z}_i = \hat{Y}_i + \hat{V}_i$, we must show $\mathbb{P}(\hat{V}_i \neq 0) \leq$ RHS(5.19). But, by Markov

$$\mathbb{P}(\hat{V}_i \neq 0) \leq \mathbb{E}[\hat{V}_i] = (w_i - \lambda_i) \leq \text{RHS}(5.19).$$

Exercise

(b) To show (b), it suffices to couple $(D_i)_{i\in[n]}$ to an indep. vector $(\hat{D}_i)_{i\in[n]}$ so that

$$\mathbb{P}[(D_i)_{i\in[n]} \neq (\hat{D}_i)_{i\in[n]}] = o(1).$$

To this end, let $(\hat{X}_{ij})_{i<j}$ be an independent copy of $(X_{ij})_{i<j}$, and set

$$\hat{D}_i = \sum_{j<i} \hat{X}_{ij} + \sum_{j=i+1}^{n} X_{ij}.$$

Then, (a) $\hat{D}_i \stackrel{d}{=} D_i$ (b) While D_i, D_j are dependent, as they both depend on X_{ij}; $\hat{D}_i, \hat{D}_j$ are independent since, for $i<j$, $\hat{D}_i$ contains $\hat{X}_{ij}$ and $\hat{D}_j$ contains X_{ij}. Finally $(D_i)_{i\in[n]} \neq (\hat{D}_i)_{i\in[n]}$ iff there is $i,j \in [n]$ with $X_{ij} \neq \hat{X}_{ij}$. Hence

$$\mathbb{P}((D_i)_n \neq (\hat{D}_i)_n) \leq 2\sum_{i,j\leq n} p_{ij}(1-p_{ij}) \leq 2\sum_{i,j\leq n} p_{ij} \xrightarrow{n\to\infty} 0 \quad \square.$$

We now investigate all degrees simultaneously. We define

(5.20) $$P_k^n = \frac{1}{n}\sum_{i\in[n]} \mathbb{1}_{\{D_i = k\}}$$

the empirical degree distribution of $GRG_n(w)$.
Due to (5.8)(5.18), we may expect that P_k^n converges to a "mixture of Poisson distributions". Formally, recall that $\mathbb{P}[\mathrm{Pois}(w_i) = k] = e^{-w_i}\frac{w_i^k}{k!}$. Thus set

(5.21) $$p_k = \mathbb{E}\left[e^{-W}\frac{W^k}{k!}\right].$$

(5.22) <u>Theorem</u>: (5.8 (a,b)). For every $\varepsilon > 0$

$$\mathbb{P}[2|p_k, P_k^n|_{TV} \geq \varepsilon] = \mathbb{P}\left[\sum_{k=0}^{\infty} |p_k - P_k^n| \geq \varepsilon\right] \xrightarrow[n\to\infty]{} 0$$

<u>Proof:</u> Since p_k is a probability distribution, it suffices to

(5.23) show $\max_{k\geq 0} |P_k^n - p_k| \xrightarrow{\mathbb{P}} 0$. To this end, observe

$$\mathbb{P}\left(\max_{k\geq 0} |P_k^n - p_k| \geq \varepsilon\right) \leq \sum_{k\geq 0} \mathbb{P}(|P_k^n - p_k| \geq \varepsilon).$$

From (5.18) we obtain (with some work)

$$E[P_k^n] = \mathbb{P}[D_U = k] \xrightarrow[n\to\infty]{} p_k \quad (\text{unif in } k)$$

so, for n large

(5.24) $$P(\max_k |P_k^n - p_k| \geq \varepsilon) \leq \sum_k P\left(|P_k^n - EP_k^n| \geq \frac{\varepsilon}{2}\right) \leq \frac{4}{\varepsilon^2} \sum_k \operatorname{Var} P_k^n .$$

Then,

$$E[(P_k^n)^2] = \frac{1}{n^2} \sum_{i,j} P(D_i = D_j = k) = $$
$$= \frac{1}{n^2} \sum_i P(D_i = k) + \frac{1}{n^2} \sum_{i \neq j} P(D_i = D_j = k)$$

So
$$\operatorname{Var} P_k^n \leq \frac{1}{n^2} \sum_i \left(P(D_i = k) - P(D_i = k)^2\right) + \frac{1}{n^2} \sum_{i\neq j} \left(P(D_i = D_j = k) - P(D_i = k) P(D_j = k)\right)$$

(5.25) $$\overset{(*)}{\leq} \frac{1}{n} + \frac{4}{n^2} \sum_{i,j} p_{ij} \left(P(D_i = k) + P(D_j = k)\right),$$

by a coupling argument (see below). Hence

$$(5.24) \leq \frac{1}{n} + \frac{1}{n^2} \sum_{i,j} p_{ij} \xrightarrow[(5.10)\,(5.11)]{n\to\infty} 0$$

which completes the proof. □

To see that (*) in (5.25) holds, we can again couple (D_i, D_j) with $(\hat{D}_i, \hat{D}_j)$ s.t. $D_i \overset{d}{=} \hat{D}_i$, $D_j \overset{d}{=} \hat{D}_j$ and $\hat{D}_i, \hat{D}_j$ are independent by taking $\hat{X}_{ij}$ be an independent copy of X_{ij} and setting

$$D_i = \sum_{k \neq i} X_{ik}, \quad D_j = \sum_{k\neq j} X_{jk}, \quad \hat{D}_i = D_i, \quad \hat{D}_j = \hat{X}_{ij} + \sum_{\substack{k: k\neq j,\\ k \neq i}} X_{jk}.$$

Now
$$P(D_i = D_j = k) - P(D_i = k) P(D_j = k) =$$
$$= P(D_i = D_j = k) - P(\hat{D}_i = k) P(\hat{D}_j = k)$$
$$= P(D_i = k, X_{ij} = 0, \hat{X}_{ij} = 1) + P(D_j = k, \hat{X}_{ij} = 0, X_{ij} = 1)$$

which implies (*) easily. □

GRG conditioned on its degrees:

We understand rather well the degree distribution of GRG. To understand more its structure we need:

(5.26) **Theorem:** Let $(d_i)_{i\in[n]}$ be a deterministic sequence of integers s.t. $\mathbb{P}[D_i = d_i \ \forall i\in[n]] > 0$. Then, conditioned on $\{D_i = d_i \ \forall i \in [n]\}$, $GRG_n(w)$ is distributed uniformly over all possible graphs whose degree sequence is $(d_i)_{i\in[n]}$.

Proof: Write $X = (X_{ij})_{i<j}$, $q_{ij} = 1-p_{ij} = \frac{l_n}{l_n + w_i w_j}$.

Then for $x = (x_{ij})_{i<j}$, $x_{ij}\in\{0,1\}$,

(5.27) $$\mathbb{P}(X = x) = \prod_{i<j} p_{ij}^{x_{ij}} q_{ij}^{1-x_{ij}}.$$

(5.28) We set $r_{ij} = \frac{p_{ij}}{q_{ij}} = \frac{w_i w_j}{l_n} = \frac{w_i}{\sqrt{l_n}}\cdot\frac{w_j}{\sqrt{l_n}} =: u_i u_j$.

Then, $r_{ij} = \frac{1-q_{ij}}{q_{ij}}$, and thus $q_{ij} = \frac{1}{1+r_{ij}} = \frac{1}{1+u_iu_j}$.

(5.29) $$\mathbb{P}(X=x) = \prod_{i<j} q_{ij}\, r_{ij}^{x_{ij}} = G(u)^{-1}\prod_{i<j}(u_iu_j)^{x_{ij}}$$

with

(5.30) $$G(u) = \prod_{i<j}(1+u_iu_j)$$

Here we define $x_{ij} = x_{ji}$ for $i>j$.

Observe now that if $X = x$, then $D_i = \sum_{j: j\neq i} x_{ij} =: d_i(x)$ Hence

(5.31) $$\mathbb{P}(X=x) = G(u)^{-1}\prod_{i\in[n]} u_i^{d_i(x)}.$$

Using this we can write, for x with $d_i(x) = d_i$

$$\mathbb{P}(X = x \mid D_i = d_i \ \forall i\in[n]) = \frac{\mathbb{P}(X = x \cap D_i = d_i \ \forall i)}{\mathbb{P}(D_i = d_i \forall i)}$$

(5.32) $$= \frac{\mathbb{P}(X=x)}{\sum_{y: d_i(y)=d_i}\mathbb{P}(X=y)} \overset{(5.31)}{=} \frac{\prod_{i\in[n]} u_i^{d_i(x)}}{\sum_{y: d_i(y)=d_i}\prod_{i\in[n]} u_i^{d_i(y)}} =$$

$$= \frac{1}{\#\{y : d_i(y) = d_i \forall i\}}. \qquad \square$$

We will explore the uniform graph with given degree sequence in the next chapter.

VI. CONFIGURATION MODEL

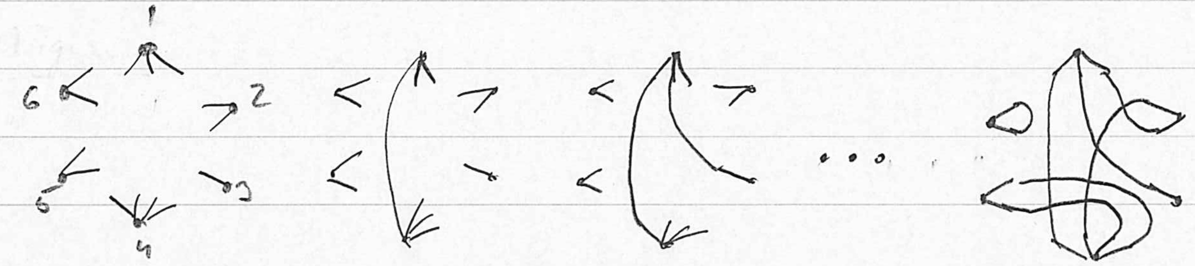
In many situations it is more suitable to fix the degrees of vertices in a random graph a priori (e.g. to avoid the degree 0 vertices). It is thus natural to consider "uniformly random graph with given degree sequence". We may ask:

- How to "generate" such graph? This is not an easy task, since in general it is difficult to even answer seemingly easier question:
- How many are such graphs? This is non-trivial even when all degrees are the same, i.e. we consider the set of all regular graphs on n vertices.
- What are the structural properties of such graph typically. This is important in applications, e.g. to test if we have a correct model.

Configuration model:

Instead of looking at "uniformly random SIMPLE graph with a given degree sequence" we start with another model, that is easier to generate.

We consider $n \in \mathbb{N}$ and sequence $(d_i)_{i\in[n]}$ such that $\sum_{i=1}^{n} d_i = \ell_n$ is even, and $d_i \in \{1,2,3,\dots\}$. To every vertex i we now attach d_i "half edges". Every halfedge will be connected to another one to form an edge in the following way: 1) Pick a not-used halfedge according to some rule and pair it with another not-used halfedge chosen uniformly at random. 2) Repeat 1) until all halfedges are paired.



(6.1) Figure: Generating of conf. model realisation for $d = (3,2,1,4,2,2)$.

Unfortunately, this construction does not ensure that the obtained graph is SIMPLE. In general, it is a multigraph, i.e. it contains multiple edges between one pair of vertices or/and self loops.

We write $CM_n(d)$ for the obtained random multigraph, and set

(6.2)
$$X_{ij} = \#\text{ edges linking } i, j \qquad i \neq j$$
$$X_{ii} = \#\text{ loops from } i \text{ to } i$$

(6.3) Then, the degree of i satisfy
$$d_i = \sum_{j \neq i} X_{ij} + 2X_{ii}.$$

One may ask if the "arbitrary rule" used in generating the multigraph influences its distribution. The answer is no. The construction above produce always what is called "uniform pairing" or "uniform matching" of the set of halfedges (identified with $\{1, \dots, l_n\}$). We first observe

(6.4) Exercise: Show that there are $(2m-1)!! = (2m-1)(2m-3)\dots 3\cdot 1$ different pairings of numbers $\{1, \dots, 2m\}$.

Now let $x_i \in \{1, \dots, l_n\}$ be the first halfedge paired in i-th step, and $y_i \in \{1, \dots, l_n\}$ the halfedge with which x_i is paired.

(6.5) Lemma: Assume that the choice of x_i depends only on $(x_j, y_j)_{j<i}$, and that for $m = 1, \dots, l_n/2$

$$P(x_m \text{ is paired to } y_m \mid x_m, (x_j, y_j)_{j<m}) = \frac{1}{l_n - 2m + 1}$$

for every $y_m \in \{x_1, \dots, x_m\} \cup \{y_1, \dots, y_{m-1}\}$. Then the resulting multigraph has the same distribution, i.e. the distribution of $CM_n(d)$.

Proof: It is sufficient to show that every pairing of $(x_i, y_i)_{i \le \frac{l_n}{2}}$ has the same probability. But

$$P(x_i \text{ is paired to } y_i \ \forall i \in [\tfrac{l_n}{2}]) =$$
$$= \prod_{m=1}^{l_n/2} P(x_m \text{ is paired to } y_m \mid x_m, (x_j, y_j)_{j<m})$$
$$= \prod_{m=1}^{l_n/2} \frac{1}{l_n - 2m + 1} = ((l_n - 1)!!)^{-1}. \qquad \square$$

With this it is not difficult to determine the law of the multigraph $CM_n(d)$.

(6.6) Proposition: (The law of $CM_n(d)$). Let $G = (x_{ij})_{i,j \in [n]}$ be a multigraph on $[n]$ (i.e. $x_{ij} = x_{ji} = \#$ of edges between i, j, $x_{ii} = \#$ loops on i) with $d_i = \sum_{j \ne i} x_{ij} + 2x_{ii}$. Then

$$P(CM_n(d) = G) = \frac{1}{(l_n - 1)!!} \cdot \frac{\prod_{i \in [n]} d_i!}{\prod_{i \in [n]} 2^{x_{ii}} \prod_{1 \le i < j \le n} x_{ij}!} \tag{6.7}$$

Proof: It is sufficient to show that the second fraction on the RHS of (6.7) is the number of pairings $N(G)$ that give the same multigraph G, (6.7) then follows from (6.4). To see this, observe that permuting the halfedges attached to given vertex does not change the graph, which explains $\prod_{i \in [n]} d_i!$ in the numerator. Some of these permutations give rise to same configurations. The factor $x_{ij}!$ compensates for multiple edges between i, j; the factor $2^{x_{ii}}$ compensates for two possible orderings of halfedges in one loop. □.

(6.8) Remark: Observe that $CM_n(d)$ is not uniform on the set of all multigraphs with given degree sequence. On the other hand, (6.6) easily imply:

(6.9) Lemma: Let $G = (x_{ij})$ be a simple graph on $[n]$ with degree sequence d (i.e $x_{ij} \in \{0,1\}$ for $i \neq j$, $x_{ii} = 0$ $i \in [n]$, $\sum_{i \neq j} x_{ij} = d_i$). Then

$$P\left(CM_n(d) = G \,\middle|\, CM_n(d) \text{ is simple} \right)$$

is independent of G. That is $CM_n(d)$ conditioned on being simple is a uniformly random simple graph with degree sequence

Assumptions on the degree sequence as $n \to \infty$

Similarly as in Chapter V, we need certain regularity on

(6.10) $$d = d_n = (d_1, \dots, d_n) = (d_1^n, \dots, d_n^n).$$

As above, let U be uniform on $[n]$ and set

(6.11) $$D_n = d_U^n = d_U$$

The distribution function of D_n is

(6.12) $$F_n(a) = \frac{1}{n} \sum_{j \in [n]} 1\!\!1_{\{d_j \le a\}}.$$

(6.13) Condition: There exists a $\{1,2,\dots\}$ valued r.v. D s.t.

(a) $$D_n \xrightarrow{d} D$$

(b) $$\lim_{n \to \infty} ED_n = ED$$

(c) $$\lim_{n \to \infty} ED_n^2 = ED^2.$$